Universidade Federal do Rio Grande do Sul

Instituto de Matemática

Programa de Pós-Graduação em Matemática

Cálculo do raio espectral de matrizes positivas e medida de Gibbs.

Dissertação de Mestrado

REGINALDO FABIANO DA SILVA AFONSO

Porto Alegre, janeiro de 2013

Dissertação submetida por Reginaldo Fabiano da Silva Afonso* , como requisito parcial para a obtenção do grau de Mestre em Ciência Matemática, pelo Programa de Pós-Graduação em Matemática, do Instituto de Matemática da Universidade Federal do Rio Grande do Sul.

Professora Orientadora:
Prof(a). Dr(a). Joana Mohr

Banca examinadora:
Prof(a). Dr(a). Joana Mohr (PPG-MAT - UFRGS)
Prof. Dr. Carlos Felipe Lardizabal Rodrigues (PPG-MAT - UFRGS)
Prof. Dr. Jairo Krás Mengue (PPG-MAT - UFRGS)
Prof(a). Dr(a). Virgínia Maria Rodrigues (PPG-MAp - UFRGS)

*Bolsista do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)

Agradecimentos

A Universidade Federal do Rio Grande do Sul, em particular, ao Instituto de Matemática - Programa de Pós-Graduação em Matemática - pela oportunidade de realização deste curso.

Ao CNPq pelo incentivo financeiro.

A minha orientadora, Prof(a) Joana, pela paciência, escolha do tema e, principalmente, pela excelente orientação.

A meus ex-colegas de mestrado, dentre os quais destaco, Danielle Azevedo e Glauber Quadros, pela disposição em compartilhar e discutir ideias.

Aos meus pais, Ivanir e Maria Eli, pelo apoio incondicional durante toda esta jornada.

A Deus por me fornecer coragem nos momentos em que pensei em fraquejar.

Resumo

Neste trabalho consideramos uma matriz positiva A. Primeiramente, mostramos o Teorema de Perron para A. Este teorema afirma que A tem um autovalor positivo, que é igual ao seu raio espectral, e associado a este autovalor temos um autovetor com entradas positivas, que chamamos de raiz de Perron e vetor de Perron de A, respectivamente. Em um segundo momento usamos análise matricial aliada a programação geométrica para descrever um método que permite o cálculo do raio espectral de uma matriz positiva. Por fim, aplicamos as ferramentas de programação geométrica na teoria de formalismo termodinâmico, no caso de um observável que depende apenas das duas primeiras coordenadas, buscando exibir a medida de Gibbs associada a este observável.

Abstract

In this work we consider a positive matrix $A_{n\times n}$. In a first moment we show Perron´s theorem for A. This theorem claims that A has a positive eigenvalue, which is equal to its spectral radius, and associated with this eigenvalue we have a eigenvector with positive entries, which we call Perron root and Perron vector of A, respectively. In a second moment we use matrix analysis and geometric programming to describe a method which allows the calculation of the spectral radius of a positive matrix. In the end, we also use the tools of geometric programming in the theory of thermodynamical formalism, in the case where we have an observable which depends only on the two first coordinates, to exhibit the Gibbs measure associated with this observable.

Índice

Introdução

A área de dinâmica simbólica é um ramo de sistemas dinâmicos que tem apresentado um rápido desenvolvimento. Embora tenha sido originado como um método para estudar sistemas dinâmicos em geral, tem mostrado significativa aplicabilidade, por exemplo no armazenamento e transmissão de dados.

A programação geométrica é outra área do conhecimento que tem ganhado importância nos últimos tempos, visto que ela permite tratar uma ampla classe de problemas, e em contrapartida, nos fornece dados quantitativamente úteis.

O presente trabalho transcorre sobre ambos os temas, de modo que os conteúdos estão dispostos da seguinte maneira: Durante o desenvolvimento do primeiro capítulo nosso foco primordial é demonstrar o teorema de Perron. O segundo capítulo trata sucintamente sobre programação geométrica. Nosso objetivo nessa altura do trabalho é preparar o terreno para que, utilizando os conceitos de programação geométrica, seja possível elaborar um método que permita o cálculo da raiz de Perron de uma matriz positiva. Este capítulo trata das relações entre os programas primal A e dual B, os quais são alguns dos pilares de programação geométrica. Visando manter o foco criamos o apêndice composto por 3 seções. Em (A.1) tratamos do lema de Farkas, em (A.2) abordamos propriedades de funções convexas em conjuntos convexos e, por fim, em (A.3) apresentamos as demonstrações dos teoremas 2.1.3 e 2.1.4 enunciados na secção (2.1). Optamos por deslocar tais demonstrações para (A.3) por se tratarem de demonstrações extensas e consideravelmente técnicas. O terceiro capítulo também está subdividido em três seções. A primeira apresenta um método o qual permite a estimativa da raiz de Perron, sendo este proveniente da programação geométrica. A segunda mostra como a programação geométrica pode ser utilizada no cálculo da raiz de Perron. Por fim, na terceira seção, reformulamos o programa dual obtido na seção anterior, e aplicamos este resultado em formalismo termodinâmico, no contexto abaixo.

Consideramos o espaço de Bernoulli dado pelas sequências $\Sigma := \{1, ..., n\}^{\mathbb{N}}$ ($x \in \Sigma$, $x = (x_1, x_2, ...)$, com $x_i \in \{1, ..., n\}$), a dinâmica dada pelo shift ($\sigma : \Sigma \to \Sigma$ definido por $\sigma(x_1, x_2, ...) = (x_2, x_3, ...)$) e $\mathcal{M}_\sigma$ as medidas invariantes pelo shift. Dada $f : \Sigma \to \mathbb{R}$, estamos interessados em encontrar uma medida que satisfaça o seguinte princípio variacional

$$P(f) = \sup_{\mu \in \mathcal{M}_\sigma} \left\{ \int f d\mu + h_\mu(\sigma) \right\}, \tag{1}$$

onde $h_\mu(\sigma)$ é a entropia da medida, $P(f)$ é chamada de pressão de f e a medida que atinge este supremo é chamada de medida de Gibbs.

No caso em que f depende apenas das duas primeiras coordenadas do espaço Σ, podemos associar a f uma matriz B $n \times n$ com entradas $b_{ij} = e^{f(i,j)}$. É conhecido que neste caso $P(f) = \log \lambda_B$, onde λ_B é o maior autovalor da matriz B. Usando o programa dual B exibimos a medida μ^* que realiza o supremo

$$P(f) = \log \lambda_B = \int f d\mu^* + h_{\mu^*}(\sigma),$$

ou seja, μ^* é a medida de Gibbs.

Capítulo 1

Pré-requisitos

Neste capítulo, revisaremos definições e resultados que serão utilizados ao longo dessa dissertação, culminando no teorema de Perron, bem como sua demonstração.

As principais referências utilizadas na elaboração deste capítulo, foram: [2], [5], [6], [7], [8] e [9].

1.1 Matrizes não-negativas: conceitos e propriedades

Começaremos com algumas definições e propriedades sobre matrizes não-negativas.

Definição 1.1.1. *Uma matriz* $A = [a_{ij}] \in M_{m\times n}(\mathbb{R})$ *é dita não-negativa, e escreve-se* $A \geq 0$*, quando* $a_{ij} \geq 0$; $\{(i = 1, ..., m),\ (j = 1, ..., n)\}$. *Similarmente uma matriz* $A = [a_{ij}] \in M_{m\times n}(\mathbb{R})$ *é dita positiva, e escreve-se* $A > 0$*, quando* $a_{ij} > 0$; $\{(i = 1, ..., m),\ (j = 1, ..., n)\}$. *De um modo geral, dadas* $A, B \in M_{m\times n}(\mathbb{R})$*, escreve-se* $A \geq B$*, quando* $a_{ij} \geq b_{ij}$; $\{(i = 1, ..., m),\ (j = 1, ..., n)\}$.

Diretamente da Definição de matrizes positivas, não-negativas e operações elementares com matrizes podemos concluir os seguintes resultados:

Sejam $P \in M_{m\times n}(\mathbb{R}), N \in M_{m\times n}(\mathbb{R})$ e $u, v, x, z \in \mathbb{R}^n$:

i- $P > 0; x \geq 0; x \neq 0 \Rightarrow Px > 0.$

ii- $N \geq 0; u \geq v \geq 0 \Rightarrow Nu \geq Nv.$

iii-$N \geq 0; z > 0; Nz = 0 \Rightarrow N = 0.$

iv- $N \geq 0; u > v > 0; N \neq 0 \Rightarrow Nu > Nv.$

Definição 1.1.2. *Seja $A = [a_{ij}] \in M_{n\times n}(\mathbb{R})$, $\lambda \in \mathbb{C}$ e $x \in \mathbb{C}^n - \{0\}$ satisfazendo a igualdade, $Ax = \lambda x$. Nestas condições, λ e x são denominados respectivamente autovalor e autovetor de A. Todo o par da forma (λ, x), que satisfaz a igualdade presente nesta definição denomina-se um autopar. Denotaremos o conjunto de autovalores distintos de A por $\sigma(A)$, o qual é chamado de espectro de A.*

Definição 1.1.3. *Seja $A = [a_{ij}] \in M_{n\times n}(\mathbb{R})$. O raio espectral de A, denotado por $\rho(A)$, é definido pela igualdade $\rho(A) = \max\limits_{\lambda\in\sigma(A)} |\lambda|$.*

Definição 1.1.4. *Seja $A = [a_{ij}] \in M_{m\times n}(\mathbb{R})$. O espaço nulo de A, denotado por $\mathcal{N}(A)$, é o conjunto definido pela igualdade $\mathcal{N}(A) = \{x \in \mathbb{C}^n | Ax = 0\}$.*

Teorema 1.1.5 (Forma canônica de Jordan)**.** *a)Seja A uma matriz complexa quadrada, $A = [a_{ij}] \in M_{n\times n}(\mathbb{C})$. Existe uma matriz inversível P tal que*

$$P^{-1}AP = J = \begin{pmatrix} J_1 & 0 & \cdots & 0 & 0 \\ 0 & J_2 & & \vdots & \vdots \\ \vdots & 0 & \ddots & 0 & \vdots \\ \vdots & \vdots & & J_{(s-1)} & 0 \\ 0 & 0 & \cdots & 0 & J_s \end{pmatrix}$$

, onde $J_l = \begin{pmatrix} \lambda_l & 1 & \cdots & 0 & 0 \\ 0 & \lambda_l & & \vdots & \vdots \\ \vdots & 0 & \ddots & 1 & \vdots \\ \vdots & \vdots & & \lambda_l & 1 \\ 0 & 0 & \cdots & 0 & \lambda_l \end{pmatrix}$ *são matrizes $r_l \times r_l$, λ_l é autovalor de A e*

$\sum_{l=1}^{s} r_l = n$, *dizemos que* J_l *é um bloco de Jordan correspondente ao autovalor* λ_l.

b) O número máximo de autovetores linearmente independentes associados a um autovalor λ_l *é igual ao número de blocos de Jordan correspondente a* λ_l. *A soma das ordens dos blocos de Jordan correspondentes a um autovalor* λ_l *é igual ao número de vezes que* λ_l *é raiz da equação característica.*

Para uma demonstração deste teorema veja [9] e [2].

Definição 1.1.6. *Seja* λ *um autovalor da matriz* A $n \times n$. *Definimos*
i) $mult\, alg_A(\lambda)$ *número de vezes que* λ *é raiz da equação característica,*
ii) $mult\, geo_A(\lambda)$ *o número máximo de autovetores linearmente independentes associados a* λ,
iii) $ind(\lambda)$, *o índice de* λ, *a dimensão do maior bloco de Jordan associado a* λ.

Observação: Segue do item (b) do teorema 1.1.5 que a multiplicidade geométrica de um autovalor λ é o número de blocos de Jordan correspondente ao autovalor λ e que a multiplicidade algébrica do autovalor λ é a soma das ordens dos blocos de Jordan correspondentes a λ, em particular se $ind(\lambda) = 1$ então $mult\, alg_A(\lambda) = mult\, geo_A(\lambda)$.

Definição 1.1.7. *Uma matriz* $A = [a_{ij}] \in M_{n\times n}(\mathbb{C})$ *é dita nilpotente, quando* $A^k = 0$, *para algum inteiro positivo* k.

Vejamos alguns lemas que irão ajudar ao longo do caminho:

Lema 1.1.8. *Se* A *é uma matriz quadrada,* $A = [a_{ij}] \in M_{n\times n}(\mathbb{R})$, $A > 0$, *então,* $\rho(A) > 0$.

Demonstração: De fato, suponhamos por absurdo que $\rho(A) = 0$. Seja B a forma de Jordan de A. Como os autovalores de A são todos nulos, segue-se que B é nilpotente. Agora, tanto A como B representam o mesmo operador linear, logo existe uma matriz P invertível, tal que, $B = P^{-1}AP$. Ora, sendo B nilpotente, existe $k \in \mathbb{N}$ tal que $B^k = 0$, donde vem que, $0 = P^{-1}A^kP$. Mas como P é invertível, segue-se que $A^k = 0$, o que é um absurdo, visto que $A > 0$.

$\blacksquare$

Lema 1.1.9. *Se A é uma matriz quadrada, $A = [a_{ij}] \in M_{n\times n}(\mathbb{R})$, $A > 0$. Temos a seguinte equivalência: $A > 0 \Leftrightarrow \frac{A}{r} > 0$, onde $\rho(A) = r > 0$. Neste caso, $\rho\left(\frac{A}{r}\right) = 1$.*

Demonstração: ($\Rightarrow$) É consequência do lema 1.1.8 e da definição de multiplicação de uma matriz por um escalar. Falta mostrar que $\rho\left(\frac{A}{r}\right) = 1$. Visando determinar os autovalores de $\frac{A}{r}$ devemos descobrir para quais escalares β a equação: $\det\left(\frac{A}{r} - \beta I\right) = 0$ é satisfeita. Ora, mas pelas propriedades de determinantes temos:

$$\det\left(\frac{A}{r} - \beta I\right) = 0 \Leftrightarrow \det\left(\frac{I}{r}\right)\det(A - \beta r I) = 0 \Leftrightarrow \det(A - \beta r I) = 0.$$

Logo, $\beta r = \lambda_i$, onde λ_i é autovalor de A. Denotemos por β_i os autovalores de $\frac{A}{r}$. Pela igualdade anterior, temos: $\beta_i = \frac{\lambda_i}{r}$. Sendo $\rho(A) = r$, segue o resultado.

($\Leftarrow$) Basta usar a definição de multiplicação de uma matriz por um escalar.

$\blacksquare$

Definição 1.1.10. *Seja V um espaço vetorial arbitrário sobre um corpo $\mathbb{K}$, ($\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$). Uma função $\|.\| : V \to \mathbb{R}$ é uma norma se as seguintes condições são satisfeitas:*

i- $\|x + y\| \leq \|x\| + \|y\|, \forall x, y \in V$

ii- $\|\alpha x\| = |\alpha| \|x\|, \forall x \in V, \forall \alpha \in \mathbb{K}$

iii- $\|x\| \geq 0$ *e* $\|x\| = 0 \Leftrightarrow x = 0$

Exemplo 1.1.11. *Dado $x \in \mathbb{C}^n$, definimos a norma da soma como $||x||_1 = \sum_{i=1}^{n} |x_i|$.*

Exemplo 1.1.12. *Dado $p \geq 1$, a **norma p** de $x \in \mathbb{C}^n$, denotada como $||x||_p$, é* *definida por* $||x||_p = \left(\sum\limits_{i=1}^{n} |x_i|^p \right)^{1/p}$. *De fato a norma* p *caracteriza uma norma em* $\mathbb{C}^n$*, pois as propriedades (i), (ii) e (iii) são satisfeitas.*

Observe que quando $p = 1$ as definições de norma soma e norma p coincidem.

Exemplo 1.1.13. *A* ***norma infinito*** *de* $x \in \mathbb{C}^n$*, denotada como* $||x||_\infty$*, é definida por* $||x||_\infty = \max\limits_{i \in \{1,...,n\}} \{|x_i|\}$.

Exemplo 1.1.14. *A* ***norma 1*** *de* $A = [a_{ij}] \in M_{m\times n}(\mathbb{C})$*, denotada como* $||A||_1$*, é definida por* $||A||_1 = \max\limits_{\|x\|_1=1} \{\|Ax\|_1\}$*. A norma 1 é um exemplo de norma em* $M_{m\times n}(\mathbb{C})$*, pois as condições (i)-(iii) presentes na definição de norma são satisfeitas.*

Exemplo 1.1.15. *A* ***norma infinito*** *de* $A = [a_{ij}] \in M_{m\times n}(\mathbb{C})$*, denotada como* $||A||_\infty$*, é definida por* $||A||_\infty = \max\limits_{\|x\|_\infty=1} \{\|Ax\|_\infty\}$.

Teorema 1.1.16. *Seja* A *uma matriz cujas entradas são números complexos,* $A = [a_{ij}] \in M_{m\times n}(\mathbb{C})$*, então,* $||A||_\infty = \max\limits_{i \in \{1,...,m\}} \left\{ \sum\limits_{j=1}^{n} |a_{ij}| \right\}$.

Demonstração: Seja $x \in \mathbb{C}^n$, um vetor qualquer de modo que $||x||_\infty = 1$. A desigualdade triangular aliada a definição de norma infinito de um vetor e norma infinito de uma matriz nos fornece a seguinte sequência de desigualdades:

$$||A||_\infty = \max_{||x||_\infty=1} ||Ax||_\infty = \max_{i \in \{1,...,m\}} \left\{ \left| \sum_{j=1}^{n} a_{ij}x_j \right| \right\} \leq \max_{i \in \{1,...,m\}} \left\{ \sum_{j=1}^{n} |a_{ij}||x_j| \right\} \leq$$
$$\leq \max_{i \in \{1,...,m\}} \left\{ \sum_{j=1}^{n} |a_{ij}| \right\}.$$

Agora, resta mostrar que o valor $\max\limits_{i \in \{1,...,m\}} \left\{ \sum\limits_{j=1}^{n} |a_{ij}| \right\}$ é atingido para algum vetor $x \in \mathbb{C}^n$ onde, $||x||_\infty = 1$.

Suponha que o máximo $\max\limits_{i \in \{1,...,m\}} \left\{ \sum\limits_{j=1}^{n} |a_{ij}| \right\}$ ocorra na linha i_0. Tome $x = (x_1, ..., x_n) \in \mathbb{C}^n$ da seguinte maneira:

$x_j = \frac{\overline{a_{i_0 j}}}{|a_{i_0 j}|}$. Claramente $\|x\|_\infty = 1$ e $\|Ax\|_\infty = \sum\limits_{j=1}^{n} |a_{i_0 j}|$

■

Em breve vamos tratar sobre convergência de sequências de matrizes. Mas ao abordarmos convergência, teremos que nos referir a uma norma específica, a menos que as normas sejam equivalentes.

De fato, a equivalência de normas de matrizes é realmente válida. Uma maneira de verificarmos isto é "imitarmos"as demonstrações apresentada em [7] nas páginas 17 e 19, pelos teoremas 5 e 8, mostrando que toda a sequência limitada de matrizes possui uma subsequência convergente, e que qualquer norma é equivalente a norma infinito.

Teorema 1.1.17. *Seja A uma matriz complexa quadrada, $A = [a_{ij}] \in M_{n\times n}(\mathbb{C})$. Nestas condições, $\lim\limits_{k\to\infty} A^k = 0$, se e somente se, $\rho(A) < 1$.*

Demonstração: Seja J a forma de Jordan da matriz A, $J = P^{-1}AP$, onde, $J = \begin{pmatrix} J_1 & 0 & \cdots & 0 & 0 \\ 0 & J_2 & & \vdots & \vdots \\ \vdots & 0 & \ddots & 0 & \vdots \\ \vdots & \vdots & & J_{(s-1)} & 0 \\ 0 & 0 & \cdots & 0 & J_s \end{pmatrix}$ de modo que as submatrizes J_l são quadradas

com a seguinte forma: $J_l = \begin{pmatrix} \lambda_l & 1 & \cdots & 0 & 0 \\ 0 & \lambda_l & & \vdots & \vdots \\ \vdots & 0 & \ddots & 1 & \vdots \\ \vdots & \vdots & & \lambda_l & 1 \\ 0 & 0 & \cdots & 0 & \lambda_l \end{pmatrix}$.

Assim, $A^k = PJ^kP^{-1}$, onde $J^k = \begin{pmatrix} {J_1}^k & 0 & \cdots & 0 & 0 \\ 0 & {J_2}^k & & \vdots & \vdots \\ \vdots & 0 & \ddots & 0 & \vdots \\ \vdots & \vdots & & {J_{(s-1)}}^k & 0 \\ 0 & 0 & \cdots & 0 & {J_s}^k \end{pmatrix}$.

Claramente, $A^k \to 0 \Leftrightarrow J^k \to 0$. Logo basta mostrar a equivalência ${J_p}^k \to 0 \Leftrightarrow |\lambda_p| < 1, \forall p \in 1, ..., s$. Vamos fazer a demonstração para $p = 1$, para os demais o procedimento é análogo.

($\Rightarrow$)Definimos $\begin{pmatrix} s \\ t \end{pmatrix} = 0$, quando $s < t$. Por indução sobre k, mostra-se que:

$$ {J_1}^k = \begin{pmatrix} \lambda_1^k & \begin{pmatrix} k \\ 1 \end{pmatrix} \lambda_1^{k-1} & \cdots & \begin{pmatrix} k \\ m-2 \end{pmatrix} \lambda_1^{k-m+2} & \begin{pmatrix} k \\ m-1 \end{pmatrix} \lambda_1^{k-m+1} \\ & \lambda_1^k & \begin{pmatrix} k \\ 1 \end{pmatrix} \lambda_1^{k-1} & \ddots & \vdots \\ & & \ddots & \ddots & \begin{pmatrix} k \\ 2 \end{pmatrix} \lambda_1^{k-2} \\ & & & \lambda_1^k & \begin{pmatrix} k \\ 1 \end{pmatrix} \lambda_1^{k-1} \\ & & & & \lambda_1^k \end{pmatrix}_{m \times m}. $$

Consequentemente, quando $k \to \infty$ temos a seguinte cadeia de implicações:

$$ {J_1}^k \to 0 \Rightarrow |\lambda_1|^k \to 0 \Rightarrow |\lambda_1| < 1 $$

($\Leftarrow$) Agora, suponha que $|\lambda_1| < 1$. Sob esta hipótese temos dois casos:

Caso 1: $\lambda_1 = 0$ - Neste caso, segue-se que J_1 é nilpotente, e consequentemente, ${J_1}^k \to 0$, quando $k \to \infty$

Caso 2: $\lambda_1 \neq 0, |\lambda_1| < 1$ Sabemos por [6] que $\lim\limits_{n\to\infty} \frac{a^n}{n^p} = \infty, p \in \mathbb{N}, a > 1$. Além disso, da desigualdade, $\begin{pmatrix} k \\ j \end{pmatrix} = \frac{k(k-1)...(k-j+1)}{j!} \leq \frac{k^j}{j!}$ segue-se que:

$$\begin{pmatrix} k \\ j \end{pmatrix} |\lambda_1^{k-j}| \leq \frac{k^j}{j!}|\lambda_1^{k-j}| = \frac{|\lambda_1|^{-j}}{j!}(k^j|\lambda_1|^k).$$

Note que, para j fixo, definindo, $a = \frac{1}{|\lambda_1|}$ e utilizando o resultado relembrado acima, $\lim\limits_{k\to\infty} \left(k^j|\lambda_1|^k\right)^{-1} = \infty$. Consequentemente, $\lim\limits_{k\to\infty} \left(k^j|\lambda_1|^k\right) = 0$. Assim utilizando o teorema do sanduiche [5], vem que $\begin{pmatrix} k \\ j \end{pmatrix} |\lambda_1^{k-j}| \to 0$ quando $k \to \infty$. Portanto, ${J_1}^k \to 0$.

Procedendo dessa maneira para cada um dos blocos de Jordan, segue-se o resultado almejado.

■

1.2 Teorema de Perron

Durante esta seção, dada uma matriz $A = [a_{ij}] \in M_{m\times n}(\mathbb{C})$ denotaremos por $|A|$, a matriz formada pelos valores absolutos das entradas de A, isto é,
$|A| = [|a_{ij}|] \in M_{m\times n}(\mathbb{C})$. Em particular, se $x \in \mathbb{R}^n$, $|x| = (|x_1|, ..., |x_n|)$.

Teorema 1.2.1. *Seja A uma matriz quadrada, $A \in M_{n\times n}(\mathbb{R})$, $A > 0$, então as seguintes afirmações são verdadeiras.*
i- $\rho(A) \in \sigma(A)$
ii- Se $Ax = \rho(A)x$, então $A|x| = \rho(A)|x|$ e $|x| > 0$.
Em outras palavras, A tem um autopar da forma $(\rho(A), v)$ com $v > 0$.

Demonstração: Pelo lema 1.1.9 basta provar o teorema no caso em que $\rho(A) = 1$.

Se (λ, x) é qualquer autopar de A tal que $|\lambda| = 1$, então, pelas hipóteses e desigualdade triangular, temos:

$$|x| = |\lambda||x| = |\lambda x| = |Ax| \leq |A||x| = A|x|.$$

Portanto:

$$|x| \leq A|x| \tag{1.1}$$

O objetivo é mostrar que a igualdade se mantém para $|\lambda| = 1$. Por conveniência, tomemos $z = A|x|$, definamos $y = z - |x|$, e observemos que a inequação (1.1), implica em $y \geq 0$. Suponhamos, por absurdo, que $y \neq 0$. Então existe algum $y_i > 0$. Da consequência (i) da Definição 1.1.1, vem que $Ay > 0$ e $z > 0$. Assim existe um número $\varepsilon > 0$ tal que $Ay > \varepsilon z$ (por exemplo, denotando Ay por d e fazendo $\varepsilon = \frac{\min_{i\in\{1,\ldots,n\}} d_i}{2\max_{i\in\{1,\ldots,n\}} z_i}$. Isto claramente é possível visto que x, y e z estão definidos acima).

Sendo $Ay > \varepsilon z$, pela definição de y e de z isto equivale a: $Az > z + \varepsilon z$, ou seja, $\frac{A}{1+\varepsilon}(z) > z$.

Assim obtivemos a desigualdade $Bz > z$, onde $B = \left(\frac{A}{1+\varepsilon}\right)$. Aplicando a consequência (iv) da Definição 1.1.1, $B^2 z > Bz > z$. Procedendo por indução, conclui-se que:

$$B^k z > z, \forall k \in \mathbb{N} \tag{1.2}$$

Por outro lado, $\rho(B) = \rho\left(\frac{A}{1+\varepsilon}\right) = \frac{1}{1+\varepsilon} < 1$. e pelo teorema 1.1.17, vem que, $\lim\limits_{k\to\infty} B^k = 0$. Consequentemente, tomando o limite em (1.2), teríamos que $0 \geq z$, o que é uma contradição. Portanto, $y = 0$ e por conseguinte, $A|x| = |x|$

$\blacksquare$

Acima, mostramos que $\rho(A)$ é autovalor de A. Nosso objetivo agora é determinar a multiplicidade de tal autovalor.

Dada uma matriz $A = [a_{ij}] \in M_{n\times n}$, consideremos a matriz de Jordan de A a qual denotaremos por J. Lembramos que, Por definição, J é uma matriz triangular superior cuja diagonal é composta pelos autovalores de A (denotados por (λ_i), $i \in \{1, ..., r\}$). Para cada $k \in \mathbb{N}$, k fixo, como $A = PJP^{-1}$, o produto de matrizes triangulares superior é uma matriz triangular superior e por propriedades de determinantes temos as seguintes igualdades:

$$0 = \det\left(A^k - \alpha I\right) = \det\left(PJ^kP^{-1} - \alpha I\right) = \det\left(PJ^kP^{-1} - P\alpha IP^{-1}\right) =$$

$$\det\left(J^k - \alpha I\right) = \prod_{i=1}^{r}\left(\lambda_i^k - \alpha\right).$$

Assim os autovalores de A^k são as soluções da equação: $\lambda_i^k - \alpha = 0$, $i \in \{1, ..., r\}$, e consequentemente, $\rho(A^k) = (\rho(A))^k$.

No teorema 1.1.17, concluímos que $\lim\limits_{k\to\infty} A^k = 0$, se e somente se, $\rho(A) < 1$. Analisaremos o que ocorre quando $\rho(A) \geq 1$, a fim de utilizar essa informação posteriormente.

Ora, consideremos a forma de Jordan de A, conforme apresentada no teorema 1.1.17. O limite $\lim\limits_{k\to\infty} A^k$ existe se e somente se $\lim\limits_{k\to\infty} J_l^k$ existe, para cada $l \in \{1, ..., s\}$. Basicamente temos 3 casos a analisar:

(Caso 1) Algum dos autovalores tem módulo maior que 1. Sem perda de generalidade, suponhamos que $|\lambda_1| > 1$. A fórmula para J_1^k foi apresentada na demonstração do teorema 1.1.17. Suponhamos, por absurdo, que λ_1^k é convergente, utilizando as propriedades de módulo, mostra-se que, $|\lambda_1|^k$ é convergente. Sendo $|\lambda_1| > 1$, temos que, $|\lambda_1|^k$ é divergente o que contradiz nossa hipótese. Logo ${\lambda_1}^k$ é divergente. Portanto, J_1^k é divergente. Assim, neste caso, A^k é divergente.

(Caso 2) $\exists \lambda_i \; i \in \{1, ..., s\}; |\lambda_i| = 1 \, e \, \lambda_i \neq 1$. Neste caso, temos $\lambda_i = e^{i\theta}$ onde, $0 < \theta < 2\pi$. Assim os termos da diagonal de J_i^k variam indefinidamente, e isto impede J_i^k (e A^k) de ter um limite.

(Caso 3)$|\lambda_i| \leq 1, \forall i \in \{1, ..., s\}$, e ainda, $|\lambda_j| = 1$ apenas quando $\lambda_j = 1$. Seja J_1 o bloco de Jordan de maior dimensão associado ao autovalor 1. Tal bloco terá a forma: $J_1 = \begin{pmatrix} 1 & 1 & \cdots & 0 & 0 \\ 0 & 1 & & \vdots & \vdots \\ \vdots & 0 & \ddots & 1 & \vdots \\ \vdots & \vdots & & 1 & 1 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}_{r \times r}$.

Definindo $\begin{pmatrix} s \\ t \end{pmatrix} = 0$, quando $s < t$, temos:

$$J_1{}^k = \begin{pmatrix} 1 & \begin{pmatrix} k \\ 1 \end{pmatrix} & \cdots & \begin{pmatrix} k \\ r-2 \end{pmatrix} & \begin{pmatrix} k \\ r-1 \end{pmatrix} \\ & 1 & \begin{pmatrix} k \\ 1 \end{pmatrix} & \ddots & \vdots \\ & & \ddots & \ddots & \begin{pmatrix} k \\ 2 \end{pmatrix} \\ & & & 1 & \begin{pmatrix} k \\ 1 \end{pmatrix} \\ & & & & 1 \end{pmatrix}_{r \times r}.$$

Mas a matriz acima terá limite, se e somente se, $r = 1$.

Os lemas 1.2.2 e 1.2.3 serão úteis na demonstração do teorema de Perron.

Lema 1.2.2. *Seja $A \in M_{m \times n}(\mathbb{C})$ e $B \in M_{n \times p}(\mathbb{C})$, então, $\|AB\|_\infty \leq \|A\|_\infty \|B\|_\infty$.*

Demonstração: De fato. Inicialmente, observe que, se $B \in M_{r \times s}(\mathbb{C})$ e se $x \in M_{s \times 1}(\mathbb{C})$ temos:

$$\|Bx\|_\infty = \max_{i \in \{1,\ldots,r\}} \left| \sum_{j=1}^{s} b_{ij} x_j \right| \leq \max_{i \in \{1,\ldots,r\}} \sum_{j=1}^{s} |b_{ij} x_j| \leq \left(\max_{i \in \{1,\ldots,r\}} \sum_{j=1}^{s} |b_{ij}| \right) \left(\max_{j \in \{1,\ldots,r\}} |x_j| \right) =$$

$$= \left(\max_{i\in\{1,...,r\}} \sum_{j=1}^{s} |b_{ij}|\right)\left(\max_{j\in\{1,...,n\}} |x_j|\right) = \|B\|_\infty \|x\|_\infty$$

Agora, sejam, $A \in M_{m\times n}(\mathbb{C})$, $B \in M_{n\times p}(\mathbb{C})$ e $x \in M_{p\times 1}(\mathbb{C})$, de modo que, $\|x\|_\infty = 1$. Utilizando a desigualdade deduzida acima, temos:

$$\|(AB)x\|_\infty = \|A(Bx)\|_\infty \leq \|A\|_\infty \|Bx\|_\infty \leq \|A\|_\infty \|B\|_\infty \|x\|_\infty = \|A\|_\infty \|B\|_\infty$$

Consequentemente, pela definição de norma infinito de uma matriz, temos que, $\|AB\|_\infty \leq \|A\|_\infty \|B\|_\infty$. ■

Lema 1.2.3. *Dados* $z_i \in \mathbb{C}, z_i \neq 0, i = 1, ..., j$, *temos:*

$$\left|\sum_{i=1}^{j} z_i\right| = \sum_{i=1}^{j} |z_i| \Leftrightarrow \textit{Para cada } i,\ 1 \leq i \leq j, \exists \alpha_i > 0; z_i = \alpha_i z_1, \forall j \in \mathbb{N}.$$

Demonstração: ($\Rightarrow$) Procedendo por indução sobre j.

i) Para j=2:

Sejam $z_1, z_2 \in \mathbb{C}$ tais que as suas formas cartesianas são: $z_1 = a_1 + b_1 i$ e $z_2 = a_2 + b_2 i$ e ainda, $|z_1 + z_2| = |z_1| + |z_2|$. Pelas propriedades de $|.|$, temos: $\sqrt{(a_1 + a_2)^2 + (b_1 + b_2)^2} = \sqrt{(a_1)^2 + (b_1)^2} + \sqrt{(a_2)^2 + (b_2)^2}$. Donde por cálculos elementares, segue-se que: $(a_1 b_2 - b_1 a_2)^2 = 0$. Por hipótese, $z_1 \neq 0$, suponha sem perda de generalidade, que $a_1 \neq 0$. Pela última igualdade escrita nesse parágrafo, temos: $b_2 = \frac{a_2}{a_1} b_1$ e evidentemente, $a_2 = \frac{a_2}{a_1} a_1$, logo existe α_2, tal que, $z_2 = \alpha_2 z_1$, $\left(\alpha_2 = \frac{a_2}{a_1}\right)$. Agora resta mostrar que α_2 é positivo. Faremos isso por absurdo. Ora, se $\alpha_2 = 0$, então $z_2 = 0$, o que contradiz nossas hipóteses. Suponhamos, por absurdo, que $\alpha_2 < 0$. $|z_1 + z_2| = |z_1| + |z_2|$ aliado a $z_2 = \alpha_2 z_1$ e a $z_1 \neq 0$ nos permite inferir que, $|1 + \alpha_2| = 1 + |\alpha_2|$. Se $-1 < \alpha_2 < 0$, então $1 + \alpha_2 = 1 - \alpha_2$, logo, $\alpha_2 = 0$ (absurdo!). Caso $\alpha_2 \leq -1$, temos, $-(1 + \alpha_2) = 1 - \alpha_2$, ou seja, $1 = -1$, o que é uma contradição. Logo, obrigatoriamente, $\alpha_2 > 0$.

ii) Suponhamos que a propriedade seja válida até um certo $j \geq 2$, isto é, $|z_1 + ... + z_j| = |z_1| + ... + |z_j|$ com z_i não nulos, implique em $z_i = \alpha_i z_1$ com $\alpha_i > 0$.

Devemos mostrar que a propriedade é válida para $j+1$. Sejam $j+1$ números complexos, tais que: $|z_1+...+z_{j+1}| = |z_1|+...+|z_{j+1}|$.

$|z_1+...+z_{j+1}| \leq |z_1+....+z_j|+|z_{j+1}| \leq |z_1|+...+|z_{j+1}| = |z_1+...+z_{j+1}|$. Logo pela sequência de desigualdades, temos que: $|z_1+....+z_j| = |z_1|+...+|z_j|$. Aplicando a hipótese de indução, temos: $z_i = \alpha_i z_1$, com $\alpha_i > 0\ \forall i \in \{1,...,j\}$. Por outro lado,

$$\left|\sum_{i=1}^{j+1} z_i\right| = \left|\left(\sum_{\substack{i=1\\ i\neq j}}^{j+1} z_i\right) + z_j\right| \leq \left|\sum_{\substack{i=1\\ i\neq j}}^{j+1} z_i\right| + |z_j| \leq \sum_{\substack{i=1\\ i\neq j}}^{j+1} |z_i| + |z_j| = \sum_{i=1}^{j+1} |z_i| = \left|\sum_{i=1}^{j+1} z_i\right|.$$

Donde vem que: $\left|\sum_{\substack{i=1\\ i\neq j}}^{j+1} z_i\right| = \sum_{\substack{i=1\\ i\neq j}}^{j+1} |z_i|$.

Aplicando a hipótese de indução, temos que, $z_{j+1} = \alpha_{j+1} z_1$.

Assim, acabamos de demonstrar por indução que:

Dados, $\{z_j\} \subset \mathbb{C}$,$z_j \neq 0$, temos:

$$\left|\sum_{i=1}^{j} z_i\right| = \sum_{i=1}^{j} |z_i| \Rightarrow \ Para\, cada\, i,\ 1 \leq i \leq j, \exists \alpha_i > 0; z_i = \alpha_i z_1, \forall j \in \mathbb{N}.$$

($\Leftarrow$) Procedendo por indução sobre j.

i) Para j=2:

Sejam $z_1, z_2 \in \mathbb{C}$ tais que $z_2 = \alpha_1 z_1$, onde $\alpha_1 > 0$. Temos:

$$|z_1+z_2| = |z_1+\alpha_1 z_1| = (|1+\alpha_1|)\,|z_1| = (1+\alpha_1)\,|z_1| = |z_1| + \alpha_1|z_1| = |z_1| + |z_2|$$

ii) Suponhamos que a propriedade seja válida para um certo $j \geq 2$, isto é, $z_i = \alpha_i z_1$ com $\alpha_i > 0$ e z_i não nulos, implique em $\left|\sum_{i=1}^{j} z_i\right| = \sum_{i=1}^{j} |z_i|$. Devemos mostrar que a propriedade é válida para $j+1$.

Sejam $j+1$ números complexos, tais que $z_i = \alpha_i z_1$, de modo que $\alpha_i > 0$ e z_i não nulos, para $i \in \{1,...,j+1\}$. Então:

$$\left|\sum_{i=1}^{j+1} z_i\right| = \left|\sum_{i=1}^{j+1} \alpha_i z_1\right| = \left|\left(\sum_{i=1}^{j} \alpha_i z_1\right) + \alpha_{j+1} z_1\right| = \left|\left(\sum_{i=1}^{j} \alpha_i\right) z_1 + \alpha_{j+1} z_1\right|.$$

Note que, tomando no caso (i), $z_1 := \left(\sum_{i=1}^{j} \alpha_i\right) z_1$ e $\alpha_1 := \frac{\alpha_{j+1}}{\sum_{i=1}^{j} \alpha_i}$, temos:

$$\left|\left(\sum_{i=1}^{j} \alpha_i\right) z_1 + \alpha_{j+1} z_1\right| = \left|\left(\sum_{i=1}^{j} \alpha_i\right) z_1\right| + \left|\alpha_{j+1} z_1\right|.$$

Aplicando a hipótese de indução, temos:

$$\left|\left(\sum_{i=1}^{j} \alpha_i\right) z_1\right| + |\alpha_{j+1} z_1| = \left|\left(\sum_{i=1}^{j} \alpha_i z_1\right)\right| + |\alpha_{j+1} z_1| = \sum_{i=1}^{j} |z_i| + |\alpha_{j+1} z_1| = \sum_{i=1}^{j+1} |z_i|.$$

■

Teorema 1.2.4. *Seja A uma matriz quadrada, $A \in M_{n\times n}(\mathbb{R})$, $A > 0$, então as seguintes afirmações são verdadeiras.*

a- $\rho(A)$ é o único autovalor na circunferência espectral.

b- índice($\rho(A)$)=1.

Demonstração: Pela equivalência demonstrada no lema 1.1.9 basta provar o teorema no caso em que $\rho(A) = 1$.

Vimos durante a demonstração do teorema 1.2.1 que se (λ, x) é um autopar de A, tal que $|\lambda| = 1$, então, $0 < |x| = A|x|$. Assim, $0 < |x_k| = (A\,|x|)_k = \sum_{j=1}^{n} a_{kj}\,|x_j|$. Por outro lado, como $Ax = \lambda x$, temos que, $|x_k| = |\lambda|\,|x_k| = |(\lambda x)_k| = \left|\sum_{j=1}^{n} a_{kj} x_j\right|$ e portanto:

$$\left|\sum_{j=1}^{n} a_{kj} x_j\right| = \sum_{j=1}^{n} a_{kj}\,|x_j| = \sum_{j=1}^{n} |a_{kj} x_j| \tag{1.3}$$

Aplicando o lema 1.2.3 à igualdade presente na equação (1.3) temos que existem $\alpha_j > 0$, tais que, $a_{k_j} x_j = \alpha_j a_{k_1} x_1$, ou equivalentemente, $x_j = \pi_j x_1$ onde,

$\pi_j = \frac{\alpha_j a_{k_1}}{a_{k_j}} > 0$. Assim, conclui-se que, se $|\lambda| = 1$, então, $x = x_1 p$ onde
$p = (1, \pi_2, ..., \pi_n)^T > 0$.

Por hipótese, x é autovetor associado a λ, de modo que, $|\lambda| = \rho(A) = 1$ e pelo teorema 1.2.1, segue-se que, $x_1 \neq 0$. Assim:

$\lambda x = Ax \Rightarrow \lambda p = Ap = |Ap| = |\lambda p| = |\lambda||p| = |p| = p$. Donde vem que, $\lambda = 1$.

Agora vamos mostrar que índice($\rho(A)$)=1.

Suponhamos, por absurdo, que índice(1)=m>1. Consideremos a forma de Jordan $J = P^{-1}AP$, sendo índice(1)=m>1, existe um bloco de Jordan J_* de dimensão $m \times m$ em J. Sem perda de generalidade, suponhamos que $J_* = J_1$. Pela reflexão que precedeu o lema 1.2.2, temos que, $||J_1^k||_\infty \to +\infty$, quando $k \to +\infty$. Assim, $||J^k||_\infty \to +\infty$, quando $k \to +\infty$. Ora, sendo $J = P^{-1}AP$, pelo lema 1.2.2, segue-se que: $||J^k||_\infty \leq ||P^{-1}||_\infty ||A^k||_\infty ||P||_\infty$, o que implicaria em: $\frac{||J^k||_\infty}{||P^{-1}||_\infty ||P||_\infty} \leq ||A^k||$, e consequentemente, $||A^k||_\infty \to \infty$ quando $k \to \infty$.

Seja A de acordo com as hipóteses do teorema, representemos A^k por
$A^k = \left[a_{ij}^{(k)}\right]$, e denotemos por i_k o índice da linha para a qual $\left\|A^k\right\|_\infty = \sum_{j=1}^{n} a_{i_k j}^{(k)}$.
Sabemos que existe uma vetor $p > 0$ tal que $p = Ap$, isto é, p é autovetor associado ao autovalor 1.

$$\|p\|_\infty \geq p_{i_k} = \sum_{j=1}^{n} a_{i_k j}^{(k)} p_j \geq \left(\sum_{j=1}^{n} a_{i_k j}^{(k)}\right)\left(\min_{i \in \{1,\ldots,n\}} p_i\right) = \left\|A^k\right\|_\infty \left(\min_{i \in \{1,\ldots,n\}} p_i\right)$$

Mas $\left\|A^k\right\|_\infty \left(\min_{i \in \{1,\ldots,n\}} p_i\right) \to \infty$ quando $k \to \infty$. Consequentemente, a desigualdade acima nos dá um absurdo, visto que p é um vetor constante. Portanto, ind(1)=1. Logo $mult\, alg_A(1) = mult\, geo_A(1)$, pela observação após a Definição 1.1.6.

$\blacksquare$

Teorema 1.2.5. *Seja A uma matriz quadrada, $A \in M_{n \times n}(\mathbb{R})$, $A > 0$, então $mult\, alg_A(\rho(A)) = mult\, geo_A(\rho(A)) = 1$*

Demonstração: Como antes, assumimos sem perda de generalidade que $\rho(A) = 1$.

Suponhamos por absurdo $mult\,geo_A(1) = m > 1$. Sejam x e y um par de autovetores independentes associados ao autovalor $\lambda = 1$, então $x \neq \alpha y$ para qualquer $\alpha \in \mathbb{C}$. Selecionemos um componente diferente de zero de y, digamos $y_i \neq 0$, e definimos $z = x - \frac{x_i}{y_i} y$. Desde $Az = z$, sabemos a partir do teorema 1.2.1 que $A|z| = |z| > 0$. Mas isto contradiz a igualdade $z_i = x_i - \frac{x_i}{y_i} y_i = 0$. Portanto, $m = 1$.

$\blacksquare$

Seja A uma matriz quadrada, $A \in M_{n\times n}(\mathbb{R})$, $A > 0$. Uma vez que $\mathcal{N}(A - \rho(A)I)$ é um espaço unidimensional que pode ser gerado por algum vetor $v > 0$. Existe um único autovetor p pertencente $\mathcal{N}(A - \rho(A)I)$ tal que $p > 0$ e $\sum_{j=1}^{n} p_j = 1$ (obtido pela normalização $p = \frac{v}{||v||_1}$). Este autovetor especial p é chamado de **vetor de Perron** de $A > 0$, e o autovalor associado $r = \rho(A)$ é chamado de **raiz de Perron de A**. Visto que $A > 0$, temos que, $A^T > 0$, e ainda, como $\rho(A) = \rho(A^T)$. Pela teoria desenvolvida até o momento, temos que existe um autopar de Perron dado por (r, p) para A, e pelo mesmo argumento, existe um autopar de Perron (r, q) para a A^T. Note que: $\left(q^T A\right)^T = A^T q = rq$, logo, $q^T A = rq^T$, onde o vetor $q^T > 0$ é chamado vetor de Perron da esquerda para a A.

Teorema 1.2.6. *Seja A uma matriz quadrada, $A \in M_{n\times n}(\mathbb{R})$, $A > 0$. Qualquer autovetor não-negativo de A é múltiplo do vetor de Perron.*

Demonstração: Seja (λ, y) um autopar para A tal que $y \geq 0$. Seja $x > 0$ o vetor de Perron para A^T, então, de acordo com a consequência (i) da Definição 1.1.1, $x^T y > 0$. Assim:

$$A^T x = \rho(A)x \Rightarrow x^T A = \rho(A)x^T \Rightarrow x^T Ay = \rho(A)x^T y \Rightarrow x^T \lambda y = \rho(A)x^T y.$$

Da última igualdade escrita acima, segue-se que, $\rho(A) = \lambda$. Sendo a multiplicidade de $\rho(A)$ igual a 1, segue-se que y e o vetor de Perron de A são LD.

$\blacksquare$

Teorema 1.2.7. *Seja A uma matriz quadrada, $A \in M_{n\times n}(\mathbb{R})$, $A > 0$. A raiz de Perron de A é dada por $r = \max\limits_{x\in N} f(x)$, onde $f(x) = \min\limits_{\substack{1\leq i\leq n\\ x_i\neq 0}} \frac{[Ax]_i}{x_i}$ e*
$N = \{x | x \geq 0,\ com\, x \neq 0\}$.

Demonstração: Se $\varepsilon = f(x)$ para $x \in N$, então $0 \leq \varepsilon x \leq Ax$. Sejam p e q^T, o vetor de Perron e o vetor a esquerda de Perron da matriz A associados com a raiz r, respectivamente. Da consequência (i) da Definição 1.1.1 $q^T x > 0$ e da consequência (ii) da mesma definição, temos:

$$\varepsilon x \leq Ax \Rightarrow q^T(\varepsilon x) \leq q^T Ax \Rightarrow \varepsilon(q^T x) \leq rq^T x \Rightarrow \varepsilon \leq r$$

Desde que, $f(p) = r$ (pois $A(p) = rp$) e $p \in N$, segue-se que $r = \max\limits_{x\in N} f(x)$.

$\blacksquare$

Juntando todos os resultados que demonstramos, temos o teorema de Perron o qual é enunciado abaixo:

Teorema 1.2.8. *Seja A uma matriz quadrada, $A \in M_{n\times n}(\mathbb{R})$, $A > 0$, com*
$r = \rho(A)$. Nestas condições as seguintes afirmações são verdadeiras.

1- $r > 0$.

*2- $r \in \sigma(A)$ (r é chamada **raiz de Perron**).*

3- $mult\, alg_A(r) = 1$.

4- Existe um autovetor $x > 0$, tal que, $Ax = rx$.

5- O vetor de Perron é o único vetor definido por:

$$Ap = rp;\ p > 0;\ ||p||_1 = 1.$$

com a exceção de múltiplos positivos de p, não existem outros autovetores não-negativos.

6- $r = \max_{x \in N} f(x)$, *onde* $f(x) = \min_{\substack{1 \leq i \leq n \\ x_i \neq 0}} \frac{[Ax]_i}{x_i}$ *e* $N = \{x | x \geq 0 \, com \, x \neq 0\}$.

Demonstração: Vide desenvolvimento da seção 1.2.

■

Capítulo 2

Programação geométrica: Propriedades pertinentes

Neste capítulo e no apêndice, revisaremos definições e resultados referentes a programação geométrica. Vamos aplicar tais resultados no desenvolvimento de um algoritmo que permite estimar a raiz de Perron e para exibir a medida de Gibbs associada a um potencial que depende apenas das duas primeiras coordenadas.

A programação geométrica foi escolhida pois permite tratar uma classe ampla de problemas, e em contrapartida é suficientemente específica de modo a se obter informações quantitativamente úteis ao processo.

As principais referências utilizadas na elaboração deste capítulo e do apêndice, foram: [1], [12], [4] e [7].

Vejamos abaixo os conceitos pertinentes ao presente trabalho.

2.1 Propriedades dos Programas primal A e dual B

Utilizaremos aqui algumas propriedades das funções convexas, tais propriedades estão enunciadas e provadas no apêndice (A.2).

Trabalhamos com espaços vetoriais N-dimensionais definidos sobre o corpo $\mathbb{R}$, denotados por E_N. Em particular, estamos interessados em funções com valores reais definidas em uma região convexa de E_N, as quais sejam compostas por N variáveis. Mais especificamente, nosso foco de estudo serão duas famílias de funções, a saber, os posinômios e as funções produto. Como estamos interessados em minimizar uma função convexa em um conjunto convexo, podemos utilizar os teoremas (A.2.1) e (A.2.7) para definir o seguinte programa convexo o qual será denominado problema 1.

Definição 2.1.1. *PROBLEMA 1- Suponha que $G_k(z)$, $k = 0, 1, ..., p$, são funções convexas em E_N. Determine um ponto z' satisfazendo as restrições*

$$G_1(z) \leq 0,\ G_2(z) \leq 0,\ ...,\ G_p(z) \leq 0$$

tal que para cada z satisfazendo as restrições, tenhamos:

$$G_0(z) \geq G_0(z').$$

Observação: Usamos o termo "programa" para designar a formulação matemática de um problema de otimização.

O programa convexo dado no problema (1) é dito ser consistente se existe pelo menos um ponto satisfazendo as restrições do mesmo. Ele é denominado superconsistente, se existe pelo menos um ponto z^* satisfazendo as restrições:

$$G_1(z^*) < 0,\ G_2(z^*) < 0,\ ...,\ G_p(z^*) < 0.$$

A cada problema (1) temos um problema (2) relacionado o qual é definido a seguir.

Definição 2.1.2. *PROBLEMA 2- Considere a função de Lagrange:*
$L : E_m \times E_p \to \mathbb{R}$ *dada por*

$$L(z, \mu) = G_0(z) + \sum_{k=1}^{p} \mu_k G_k(z)$$

onde $G_k(z)$, $k = 0, 1, 2, ..., p$, *é uma função convexa em* E_m. *Determine um ponto* (z', μ') *tal que:*

$$(i)\ \mu'_k \geq 0 \qquad\qquad k = 1, ..., p$$

$$(ii)\ \mu'_k G_k(z') = 0 \qquad\qquad k = 1, ..., p$$

e

$$(iii) L(z', \mu) \leq L(z', \mu') \leq L(z, \mu'),$$

para todo $z \in E_m$ *e para todo* $\mu_k \geq 0$.

Observação: De modo a tornar a leitura mais fluída, as demonstrações dos dois próximos resultados foram colocadas no apêndice, uma vez que são longas e bastante técnicas.

Teorema 2.1.3. *Consideremos um programa convexo definido no problema (1) superconsistente e suponha que* $G_k(z)$, $k = 0, 1, 2, ..., p$, *tem derivadas parciais contínuas em* E_M. *Então* z' *é uma solução do problema (1) se, e somente se, existe um vetor* μ', *tal que* (z', μ') *resolvem o problema (2).*

Demonstração: Vide apêndice (A.3).

Agora vamos enunciar o seguinte teorema o qual resultará na conhecida desigualdade geométrica.

Teorema 2.1.4. *Suponha que x é um vetor arbitrário com N componentes e seja δ um vetor arbitrário com N componentes não negativas. Então, estes dois vetores satisfazem a desigualdade.*

$$\sum_{i=1}^{N} x_i\delta_i \leq \left(\sum_{i=1}^{N}\delta_i\right)\log\left(\sum_{i=1}^{N}e^{x_i}\right)+\sum_{i=1}^{N}\delta_i \log\delta_i - \left(\sum_{i=1}^{N}\delta_i\right)\log\left(\sum_{i=1}^{N}\delta_i\right),$$

onde definimos $\delta log \delta$ é igual a zero, quando δ é zero. Além disso, e desigualdade torna-se uma igualdade, se, só se, existe algum número não-negativo B tal que

$$\delta_i = Be^{x_i}, \quad i = 1, 2, ..., N.$$

Demonstração: Vide apêndice (A.3).

A desigualdade geométrica, é um caso particular do teorema enunciado acima como veremos abaixo.

Considere δ como sendo um vetor positivo, o qual satisfaz a condição de normalidade, isto é,

$$\sum_{i=1}^{N}\delta_i = 1. \tag{2.1}$$

Sob essa condição, o teorema anterior resulta na seguinte desigualdade:

$$\sum_{i=1}^{N}[x_i - \log\delta_i]\delta_i \leq \log\left(\sum_{i=1}^{N}e^{x_i}\right) \tag{2.2}$$

Desde que x é arbitrário, e δ é positivo, podemos escolher,

$$x_i = \log(\delta_i T_i); \;\; i = 1, ..., N, \tag{2.3}$$

onde T_i é um número positivo arbitrário, porém fixo. Assim (2.2) resulta em:

$$\sum_{i=1}^{N}\delta_i \log T_i \leq \log\left(\sum_{i=1}^{N}\delta_i T_i\right). \tag{2.4}$$

De acordo com o teorema 2.1.4 a igualdade ocorre, se e somente se, $\delta_i = Be^{x_i} = Be^{\log(\delta_i T_i)} = B\delta_i T_i$; $i = 1, 2, ..., N$, para algum B positivo. Logo, utilizando propriedades da função logarítmica e o fato de que a função exponencial é crescente, temos:

$$\prod_{i=1}^{N}\left[T_i^{\delta_i}\right] \leq \sum_{i=1}^{N}\delta_i T_i \tag{2.5}$$

com a igualdade ocorrendo, se e somente se,

$$\exists B \in \mathbb{R}_{+}^{*}; \delta_i = B\delta_i T_i; i = 1, 2, ..., N. \tag{2.6}$$

A desigualdade (2.5) é conhecida como desigualdade geométrica. Mostramos que ela é válida quando a condição de normalidade (2.1) é satisfeita. Segue-se de (2.5) e (2.6), que a igualdade ocorre, se e somente se,

$$\exists T > 0; T_i = T;\ i = 1, ..., N.$$

Os escalares $\delta_i; i = 1, 2, ..., N$, são denominados pesos.

Definição 2.1.5. *Considere uma função*

$$g : X \to \mathbb{R}, X = \{(t_1, ..., t_m) \in \mathbb{R}^m |\ t_j > 0, \forall j \in \{1, ..., m\}\}, g = \sum_{i=1}^{n} u_i,$$

de modo que, $u_i = c_i \prod_{j=1}^{m} t_j^{a_{ij}}$*, com* c_i*,* $a_{ij} \in \mathbb{R}$ *constantes reais, sendo* $c_i > 0, i \in \{1, ..., n\}$*. Denominamos* g *definida nestas condições por posinômio.*

Exemplo 2.1.6. *A função,*

$$g : X \to \mathbb{R}, X = \{(t_1, t_2, t_3) \in \mathbb{R}^m |\ t_j > 0, \forall j \in \{1, 2, 3\}\},$$

dada por $g(t_1, t_2, t_3) = \frac{4}{3}t_1{}^{-1}t_2 + 7t_1t_2t_3{}^{\frac{-1}{3}}$ *é um exemplo de um posinômio.*

Definição 2.1.7. *Programa Primal A: Visamos determinar o valor mínimo de uma função* $g_0(t)$ *sujeito as seguintes restrições:*

$$t_1 > 0,\ t_2 > 0,\ ..., t_m > 0 \tag{2.7}$$

e

$$g_1(t) \leq 1,\ g_2(t) \leq 1,\ ...,\ g_p(t) \leq 1 \tag{2.8}$$

onde

$$g_k(t) = \sum_{i \in J[k]} c_i t_1^{a_{i1}} t_2^{a_{i2}} ... t_m^{a_{im}},\ \ k = 0, 1, ..., p\ , \tag{2.9}$$

$$J[k] = \{m_k, m_k + 1, m_k + 2, ..., n_k\},\ \ k = 0, 1, ..., p \tag{2.10}$$

e

$$m_0 = 1,\ m_1 = n_0 + 1, m_2 = n_1 + 1, ..., m_p = n_{p-1} + 1,\ n_p = n\ , \tag{2.11}$$

sendo $a_{ij} \in \mathbb{R}$, $c_i > 0$ *constantes fixadas e* m, p, n_k, m_k, n *números naturais fixados. Note que as funções* $g_k(t)$ *são posinômios.*

Denominações: O posinômio a ser minimizado, ou seja, $g_0(t)$ é conhecido como **função primal** e as variáveis $t_1, ..., t_m$ são chamadas de **variáveis primais**. As restrições impostas em (2.7) são **restrições naturais**, enquanto que as restrições apresentadas em (2.8) são **restrições forçadas**. Coletivamente estas restrições formam as **restrições primais**, as constantes a_{ij} formam uma matriz (a_{ij}) que é chamada **matriz dos expoentes**, a qual apresenta n linhas e m colunas, note que m é número de variáveis do problema e n é o número total de parcelas da forma $c_i t_1^{a_{i1}} t_2^{a_{i2}} ... t_m^{a_{im}}$, com $i \in J[k]$ e $k = 0, 1, ..., p$.

Exemplo 2.1.8. *Sejam,* g_0, g_1, g_2, *posinômios de 3 variáveis, cujas leis são* $g_0(t_1, t_2, t_3) = 3t_1$, $g_1(t_1, t_2, t_3) = 20{t_1}^{-1}{t_2}^{\frac{1}{2}} + 30 t_1 t_2 {t_3}^{-\frac{1}{3}}$, *e* $g_2(t_1, t_2, t_3) = \frac{4}{3} t_1 {t_2}^{-1}$. *O programa,*

"Determine o valor mínimo de uma função $g_0(t_1, t_2, t_3) = 3t_1$ *sujeito as seguintes restrições:*

$$t_1 > 0,\ t_2 > 0,\ t_3 > 0$$

e

$$g_1(t_1, t_2, t_3) \leq 1,\ g_2(t_1, t_2, t_3) \leq 1,"$$

é um exemplo de programa primal A.

Em tal exemplo $m = 3$, $n = 4$, $c_1 = 3, c_2 = 20, c_3 = 30, c_4 = \frac{4}{3}$, $J[0] = \{1\}$, $J[1] = \{2,3\}, J[2] = \{4\}$ *e a matriz de expoentes deste programa é,*

$$A = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1/2 & 0 \\ 1 & 1 & -1/3 \\ 1 & -1 & 0 \end{bmatrix}.$$

A cada programa primal A temos um programa dual B correspondente o qual é definido abaixo:

Definição 2.1.9. *Programa Dual B: Visamos determinar o valor máximo de uma função*

$$v(\delta) = \left[\prod_{i=1}^{n} \left(\frac{c_i}{\delta_i}\right)^{\delta_i}\right] \prod_{k=1}^{p} \lambda_k(\delta)^{\lambda_k(\delta)}, \tag{2.12}$$

onde

$$\lambda_k(\delta) = \sum_{i \in J[k]} \delta_i, \ k = 1, ..., p, \tag{2.13}$$

$$J[k] = \{m_k, m_k + 1, m_k + 2, ..., n_k\}, \ k = 0, 1, ..., p \tag{2.14}$$

e

$$m_0 = 1, \, m_1 = n_0 + 1, m_2 = n_1 + 1, ..., m_p = n_{p-1} + 1, \, n_p = n. \tag{2.15}$$

Os fatores c_i*, dados em (2.9), são positivos e o vetor variável* $\delta = (\delta_1, \delta_2, ..., \delta_n)$ *obedece as seguintes restrições:*

$$\delta_1 \geq 0, \, \delta_2 \geq 0, \, ..., \delta_n \geq 0 \tag{2.16}$$

$$\sum_{i \in J[0]} \delta_i = 1 \tag{2.17}$$

e

$$\sum_{i=1}^{n} a_{ij} \delta_i = 0, \quad j = 1, 2, ..., m \tag{2.18}$$

onde os coeficientes $a_{ij} \in \mathbb{R}$ *são os expoentes dados em (2.9).*

Ao avaliarmos a função produto $v(\delta)$, definimos $x^x = x^{-x} = 1$, para $x = 0$.

Denominações: A função produto, $v(\delta)$ é conhecido como **função Dual** e as variáveis $\delta_1, ..., \delta_n$ são chamadas de **variáveis duais**. As relações (2.16) são **condições de positividade**, a relação (2.17) é **condição de normalidade**, enquanto que a relação (2.18) é **condição de ortogonalidade**. Coletivamente estas condições formam as **restrições duais**.

Exemplo 2.1.10. *O programa dual B relacionado ao programa primal A, dado no exemplo 2.1.8, é apresentado pelo seguinte enunciado.*

"Determine o valor máximo da função

$$v(\delta_1, \delta_2, \delta_3, \delta_4) = \left(\frac{3}{\delta_1}\right)^{\delta_1} \left(\frac{20}{\delta_2}\right)^{\delta_2} \left(\frac{30}{\delta_3}\right)^{\delta_3} \left(\frac{\frac{4}{3}}{\delta_4}\right)^{\delta_4} (\delta_2 + \delta_3)^{(\delta_2+\delta_3)} {\delta_4}^{\delta_4}$$

sujeito as seguintes restrições:

$$\delta_1 \geq 0,\ \delta_2 \geq 0,\ \delta_3 \geq 0,\ \delta_4 \geq 0$$

$$\delta_1 = 1$$

e

$$\begin{cases} \delta_1 - \delta_2 + \delta_3 + \delta_4 = 0 \\ \frac{\delta_2}{2} + \delta_3 - \delta_4 = 0 \\ -\frac{\delta_3}{3} = 0 \end{cases}$$

Além do programa dual B, o qual está relacionado com o programa primal A, temos o programa primal transformado, que é concebido da seguinte maneira.

Consideremos o posinômio

$$g(t) = \sum_i c_i t_1^{a_{i1}} t_2^{a_{i2}} ... t_m^{a_{im}}. \tag{2.19}$$

Podemos fazer a mudança de variável,

$$t_j = e^{z_j}, \quad j = 1, 2, ...m, \tag{2.20}$$

transformando assim o posinômio anterior em

$$g(z) = \sum_i c_i e^{\sum\limits_{j=1}^{m} a_{ij} z_j}. \tag{2.21}$$

Observe que as variáveis da transformação primal z_j abrangem todos os números reais, visto que as variáveis primais t_j abrangem todos os números reais positivos. Assim, o programa primal transformado é definida da seguinte maneira:

Definição 2.1.11. *Programa Primal transformado A_z: Visamos determinar o valor mínimo de uma função $g_0(z)$ sujeito as seguintes restrições:*

$$g_1(z) \leq 1,\ g_2(z) \leq 1,\ ..., g_p(z) \leq 1, \tag{2.22}$$

onde,

$$g_k(z) = \sum_{i \in J[k]} c_i e^{\sum\limits_{j=1}^{m} a_{ij} z_j}, \quad k = 0, 1, 2, ..., p\,, \tag{2.23}$$

$$J[k] = \{m_k, m_k + 1, m_k + 2, ..., n_k\},\ k = 0, 1, ..., p \tag{2.24}$$

e

$$m_0 = 1,\ m_1 = n_0 + 1, m_2 = n_1 + 1, ..., m_p = n_{p-1} + 1,\ n_p = n \tag{2.25}$$

sendo $a_{ij} \in \mathbb{R}$, $c_i > 0$ constantes fixadas e m, p, n_k, m_k, n números naturais fixados, no programa primal A.

O leitor interessado poderá encontrar a definição do programa dual transformado na referência [1]. Omitiremos aqui tal definição, visto que não será utilizada no decorrer do trabalho.

Vejamos a seguir dois resultados os quais relacionam o programa primal A e o programa dual B. Começaremos pelo lema principal da programação geométrica.

Lema 2.1.12. *Se t satisfaz as restrições do programa primal A e δ cumpre as restrições do programa dual B, então,*

$$v(\delta) \leq g_0(t). \tag{2.26}$$

Além disso, nas condições acima, $v(\delta) = g_0(t)$, se e somente se,

$$\delta_i = \begin{cases} \frac{c_i t_1^{a_{i1}} t_2^{a_{i2}} \ldots t_m^{a_{im}}}{g_0(t)}, & i \in J[0] \\ \lambda_k(\delta) c_i t_1^{a_{i1}} t_2^{a_{i2}} \ldots t_m^{a_{im}}, & i \in J[k]; \quad k = 1, 2, \ldots, p \end{cases} \tag{2.27}$$

Demonstração: Seja t um vetor que satisfaz o programa primal A, e δ um vetor que satisfaz o programa dual B relacionado. Para cada termo $u_i(t)$ do posinômio $g_k(t)$ podemos definir:

$$x_i = \log u_i(t); \;\; i \in J[k] \tag{2.28}$$

e aplicando o teorema 2.1.4, e (2.28) obtemos,

$$\sum_{J[k]} \log u_i(t)\delta_i \leq \left(\sum_{J[k]} \delta_i\right) \log\left(g_k(t)\right) + \sum_{J[k]} \delta_i \log\delta_i - \left(\sum_{J[k]} \delta_i\right) \log\left(\sum_{J[k]} \delta_i\right),$$

o que é equivalente a,

$$\sum_{J[k]} [\log u_i(t) - \log(\delta_i)]\delta_i + \left(\sum_{J[k]} \delta_i\right) \log\left(\sum_{J[k]} \delta_i\right) \leq \left(\sum_{J[k]} \delta_i\right) \log\left(g_k(t)\right), \tag{2.29}$$

Utilizando as propriedades da função logarítmica e exponencial, temos que,

$$\left\{\prod_{J[k]} \left(\frac{u_i(t)}{\delta_i}\right)^{\delta_i}\right\} \lambda_k(\delta)^{\lambda_k(\delta)} \leq g_k(t)^{\lambda_k(\delta)} \tag{2.30}$$

onde $\lambda_k(\delta) = \sum_{J[k]} \delta_i$. Em particular, aplicando a condição de normalidade, temos,

$$\prod_{J[0]} \left(\frac{u_i(t)}{\delta_i}\right)^{\delta_i} \leq g_0(t). \tag{2.31}$$

Como consequência de (2.30), e valendo-se de que as restrições forçadas do programa primal A são satisfeitas, temos as seguintes desigualdades:

$$\left\{\prod_{J[k]} \left(\frac{u_i(t)}{\delta_i}\right)^{\delta_i}\right\} \lambda_k(\delta)^{\lambda_k(\delta)} \leq g_k(t)^{\lambda_k(\delta)} \leq 1, \quad k = 1, 2, \ldots, p \tag{2.32}$$

Das desigualdades (2.31) e (2.32), temos:

$$\left\{\prod_{J[0]}\left(\frac{u_i(t)}{\delta_i}\right)^{\delta_i}\right\}\left\{\prod_{k=1}^{p}\left\{\left\{\prod_{J[k]}\left(\frac{u_i(t)}{\delta_i}\right)^{\delta_i}\right\}\lambda_k(\delta)^{\lambda_k(\delta)}\right\}\right\} \leq g_0(t).$$

O que pode ser reescrito como:

$$\left[\prod_{i=1}^{n}\left(\frac{u_i(t)}{\delta_i}\right)^{\delta_i}\right]\left[\prod_{k=1}^{p}\lambda_k(\delta)^{\lambda_k(\delta)}\right] \leq g_0(t) \tag{2.33}$$

Agora, como $u_i(t)$ é o i-ésimo termo do programa primal A, nós temos que, $\left(\frac{u_i(t)}{\delta_i}\right)^{\delta_i} = \left(\frac{c_i}{\delta_i}\right)^{\delta_i} t_1^{a_{i1}\delta_i} t_2^{a_{i2}\delta_i} ... t_m^{a_{im}\delta_i}$. Assim, substituindo tal expressão em (2.33), segue-se que:

$$v(\delta) t_1^{\sum\limits_{i=1}^{n} a_{i1}\delta i} t_2^{\sum\limits_{i=1}^{n} a_{i2}\delta i} ... t_m^{\sum\limits_{i=1}^{n} a_{im}\delta i} \leq g_0(t). \tag{2.34}$$

Lembrando-se que por hipótese, δ satisfaz as restrições do programa dual B, em particular são válidas as condições de ortogonalidade, isto é, $\sum\limits_{i=1}^{n} a_{ij}\delta_i = 0$; $j = \{1, 2, ..., m\}$. Assim, a desigualdade (2.34) reduz-se a:

$$v(\delta) \leq g_0(t).$$

Desde modo, acabamos de deduzir a primeira conclusão do lema 2.1.12. Agora, vamos deduzir a segunda conclusão:

($\Rightarrow$) Suponhamos que $g_0(t) = v(\delta)$, onde t e δ satisfazem as restrições do programa primal A e dual B relacionado, respectivamente. Então a desigualdade (2.31) e todas as desigualdades presentes em (2.32) são igualdades. Assim,
$g_k(t)^{\lambda_k(\delta)} = 1, \quad k = 1, ..., p$ e ainda,

$$\left\{\prod_{J[k]}\left(\frac{u_i(t)}{\delta_i}\right)^{\delta_i}\right\}\lambda_k(\delta)^{\lambda_k(\delta)} = g_k(t)^{\lambda_k(\delta)}, \quad k = 0, 1, ..., p. \tag{2.35}$$

Utilizando o mesmo raciocínio empregado em (2.29) e (2.30). Temos que a igualdade acima ocorre, se e somente se,

$$\sum_{J[k]} \log u_i(t)\delta_i = \left(\sum_{J[k]}\delta_i\right)\log\left(g_k(t)\right) + \sum_{J[k]}\delta_i \log\delta_i - \left(\sum_{J[k]}\delta_i\right)\log\left(\sum_{J[k]}\delta_i\right)$$

Aplicando a segunda conclusão do teorema 2.1.4, temos que existe um número B_k, $k = 0, 1, ..., p$ tal que,

$$\delta_i = B_k u_i(t), \quad i \in J[k], \quad k = 0, 1, ..., p. \tag{2.36}$$

Somando (2.36) nos índices $i \in J[k]$, temos:

$$\lambda_k = B_k g_k(t), \quad k = 0, 1, ..., p. \tag{2.37}$$

Segue-se da condição de normalidade que,

$$B_0 = \frac{1}{g_0(t)}. \tag{2.38}$$

As equações (2.36) e (2.38) mostram a validade da relação (2.27) para $k = 0$. Vejamos a seguir por que a relação (2.27) é válida integralmente.

Ora, se $B_k = 0$, então a relação (2.37) resulta em $\lambda_k = 0$, o que significa que, $\delta_i = 0$, para todo $i \in J[k]$. Logo a relação (2.27) está justificada neste caso. Agora, analisaremos o caso em que $B_k > 0$. Neste caso, de (2.37) decorre que $\lambda_k(\delta) > 0$ e de $g_k(t)^{\lambda_k(\delta)} = 1$, segue-se que, $g_k(t) = 1$. Assim, novamente por (2.37) vem que $\lambda_k(\delta) = B_k$ e nós concluímos de (2.36) que (2.27) é satisfeita.

($\Leftarrow$) Suponhamos que t e δ satisfazem (2.27) e as restrições dos programas primal A e dual B relacionados, respectivamente. Nestas condições a segunda conclusão do teorema 2.1.4 implica que as desigualdades (2.30) e (2.31) são igualdades.

Além disso, afirmamos que as desigualdades presentes em (2.32) também são igualdades quando t e δ obedecem as condições descritas acima. De fato, se $\lambda_k(\delta) = 0$, então $\delta_i = 0$ para $i \in J[k]$ e as desigualdades (2.32) tornam-se igualdades. Resta a possibilidade em que, $\lambda_k(\delta) > 0$. Neste caso a relação (2.27), implica em: $g_k(t) = \sum_{i \in J[k]} u_i(t) = \sum_{i \in J[k]} \frac{\delta_i}{\lambda_k} =1$, e como em (2.30) temos igualdade,

$$\left\{ \prod_{J[k]} \left(\frac{u_i(t)}{\delta_i} \right)^{\delta_i} \right\} \lambda_k(\delta)^{\lambda_k(\delta)} = g_k(t)^{\lambda_k(\delta)} = 1.$$

Assim, a desigualdade (2.32) torna-se uma igualdade. Consequentemente, temos também igualdade em (2.34), e portanto, $g_0(t) = v(\delta)$.

$\blacksquare$

Teorema 2.1.13. *Suponhamos que o programa primal A é superconsistente e que a função primal $g_0(t)$ atinge seu valor mínimo em um ponto o qual satisfaz as restrições primais. Então:*

i - O programa dual correspondente B é consistente e a função dual $v(\delta)$ atinge seu valor máximo em um ponto cujo restrições duais são satisfeitas.

ii- O valor máximo da função dual é igual ao valor mínimo da função primal A.

iii- Se t' é um ponto de minimização para o programa primal A, então existem multiplicadores de Lagrange não-negativos μ'_k, $\quad k = 1, ..., p$, tais que a função de Lagrange

$$L(t,\mu) = g_0(t) + \sum_{k=1}^{p} \mu_k \left[g_k(t) - 1\right]$$

tem a propriedade

$$L(t',\mu) \leq g_0(t') = L(t',\mu') \leq L(t,\mu')$$

para arbitrários $t_j > 0$ e arbitrários $\mu_k \geq 0$. Além disso, existe um ponto de maximização δ' para o programa dual B, cujas as componentes são:

$$\delta'_i = \begin{cases} \frac{c_i t_1^{a_{i1}} t_2^{a_{i2}} ... t_m^{a_{im}}}{g_0(t)}, & i \in J[0] \\ \frac{\mu_k c_i t_1^{a_{i1}} t_2^{a_{i2}} ... t_m^{a_{im}}}{g_0(t)}, & i \in J[k]; \quad k = 1, 2, ..., p \end{cases}$$

onde $t = t'$ e $\mu = \mu'$. Além disso,

$$\lambda_k(\delta') = \frac{\mu'_k}{g_0(t')}, \quad k = 1, 2, ..., p.$$

iv- Se δ' é um ponto de maximização do programa dual B, cada ponto de minimização t' do programa primal A satisfaz o seguinte sistema de equações:

$$c_i t_1^{a_{i1}} t_2^{a_{i2}} ... t_m^{a_{im}} = \begin{cases} \delta'_i v(\delta'), & i \in J[0]; \\ \frac{\delta'_i}{\lambda_k(\delta')}, & i \in J[k]; \end{cases}$$

onde k abrange todos os inteiros positivos, para os quais $\lambda_k(\delta') > 0$.

Demonstração: Considere a função, $f : \mathbb{R}^m \to \mathbb{R}$, dada por $f(z) = \sum\limits_{j=1}^{m} a_{ij} z_j$. Pelo teorema A.2.3 f é convexa. Sendo a função exponencial crescente, podemos aplicar o teorema A.2.5, e temos que

$$P(z) = e^{\sum\limits_{j=1}^{m} a_{ij} z_j}$$

é convexa em todo $\mathbb{R}^m$. Assim, aplicando o teorema A.2.6, temos que a função

$$g(z) = \sum_{i=1}^{N} c_i e^{\sum\limits_{j=1}^{m} a_{ij} z_j}$$

é uma função convexa quando os coeficientes c_i; $i = 1, ..., N$ são constantes positivas. Assim, cada função $g_k(z), k = 0, 1, ..., p$, que compõe o programa primal transformado A_z dada pela expressão (2.23) é uma função convexa.

Agora, suponhamos que o programa primal A é superconsistente e que a função primal $g_0(t)$ apresenta um ponto de minimização t' que satisfaz as restrições primais. Então a transformação do programa primal A_z é superconsistente, e a transformação da função primal $g_0(z)$ tem um ponto de minimização z', onde,

$$z'_j = \log t'_j, \quad j = 1, 2, ..., m. \tag{2.39}$$

e

$$g_k(z') \leq 1, \quad k = 1, 2, ..., p. \tag{2.40}$$

Definindo,

$$G_k(z) = \begin{cases} g_k(z); & k = 0 \\ g_k(z) - 1; & k = 1, 2, ..., p \end{cases} \tag{2.41}$$

e aplicando o teorema 2.1.3 nós temos garantida a existência de multiplicadores de Lagrange não negativos μ'_k, $k = 1, 2, ..., p$, tais que a função de Lagrange

$$L(z, \mu) = g_0(z) + \sum_{k=1}^{p} \mu_k \left[g_k(z) - 1 \right] \tag{2.42}$$

tem a propriedade:

$$L(z',\mu) \leq g_0(z') = L(z',\mu') \leq L(z,\mu') \tag{2.43}$$

para cada z e $\mu_k \geq 0$ arbitrários. Portanto z' é um ponto de minimização para a função $L(z,\mu')$ e temos:

$$\frac{\partial L}{\partial z_q}(z',\mu') = 0, \qquad q = 1,2,...,m.$$

O que é equivalente a:

$$\sum_{i\in J[0]} a_{iq} c_i e^{\sum\limits_{j=1}^{m} a_{ij} z'_j} + \sum_{k=1}^{p} \mu'_k \left(\sum_{i\in J[k]} a_{iq} c_i e^{\sum\limits_{j=1}^{m} a_{ij} z'_j} \right) = 0, \qquad q = 1,2,...,m. \tag{2.44}$$

Dividindo ambos os lados da equação (2.44) por $g_0(z')$, temos que o vetor δ' cujas componentes são:

$$\delta'_i = \begin{cases} \dfrac{c_i e^{\sum\limits_{j=1}^{m} a_{ij} z'_j}}{g_0(z')}, & i \in J[0] \\ \dfrac{\mu'_k c_i e^{\sum\limits_{j=1}^{m} a_{ij} z'_j}}{g_0(z')} & i \in J[k], \quad k = 1,2,...,p \end{cases} \tag{2.45}$$

Notamos que este vetor satisfaz as condição de ortogonalidade no programa dual B (vide Definição 2.1.9). Além disso, por (2.45) e (2.23) constatamos que δ' definido acima obedece a condição de normalidade. Notamos também que δ' também satisfaz a condição de positividade, pois $\mu'_k \geq 0$. Desta forma, concluímos que o programa dual B é consistente.

Usando a relação (2.45) e (2.23), constatamos que,

$$\sum_{i\in J[k]} \delta'_i = \lambda_k(\delta') = \frac{\mu'_k g_k(z')}{g_0(z')}; \quad k = 1,2,...,p. \tag{2.46}$$

Então, segue-se da igualdade presente em (2.43) que, $\mu'_k\,[g_k(z') - 1] = 0$, ou seja, $\mu'_k g_k(z') = \mu'_k,\ k = 1,2,...,p$.

Assim, de (2.46), segue-se que, $\lambda_k(\delta') = \frac{\mu'_k}{g_0(z')},\ k = 1,2,...,p$.

Desta forma a relação (2.45) pode ser reescrita da seguinte maneira:

$$\delta'_i = \begin{cases} \frac{c_i e^{\sum\limits_{j=1}^{m} a_{ij} z'_j}}{g_0(z')}, & i \in J[0] \\ \lambda_k(\delta') c_i e^{\sum\limits_{j=1}^{m} a_{ij} z'_j} & i \in J[k], \quad k = 1, 2, ..., p \end{cases} \tag{2.47}$$

Agora, usando (2.39), podemos reescrever a relação acima como:

$$\delta'_i = \begin{cases} \frac{c_i t'^{a_{i1}}_1 t'^{a_{i2}}_2 ... t'^{a_{im}}_m}{g_0(t')}, & i \in J[0] \\ \lambda_k(\delta) c_i t'^{a_{i1}}_1 t'^{a_{i2}}_2 ... t'^{a_{im}}_m, & i \in J[k]; \quad k = 1, 2, ..., p \end{cases} \tag{2.48}$$

Portanto, aplicando o lema 2.1.12, inferimos que, $g_0(t') = v(\delta')$. Desta forma, pelo que fizemos até o momento, provamos as afirmações (i), (ii) e (iii) presentes no enunciado do teorema.

Resta mostrar a afirmação (iv).

Suponhamos que t' é um ponto de mínimo do programa primal A. Consideremos um ponto δ", o qual é um ponto de máximo do programa dual B. Pelo que fizemos anteriormente, vimos que existe um ponto δ' o qual é ponto de máximo para v, de modo que, $g_0(t') = v(\delta')$. Como δ" também é ponto de máximo para v, segue-se que, $g_0(t') = v(\delta")$. Consequentemente, usando a segunda conclusão do lema 2.1.12, nós inferimos que t' e δ" satisfazem a relação (2.27) e portanto,

$$c_i t'^{a_{i1}}_1 t'^{a_{i2}}_2 ... t'^{a_{im}}_m = \begin{cases} \delta''_i v(\delta''), & i \in J[0] \\ \frac{\delta''_i}{\lambda_k(\delta'')}, & i \in J[k] \end{cases}$$

onde k engloba todos os inteiros para os quais $\lambda_k(\delta") > 0$.

■

Capítulo 3

Programação geométrica e a raiz de Perron

Neste capítulo, apresentaremos um método o qual permite a estimativa da raiz de Perron de uma matriz positiva baseado nas idéias de programação geométrica. Mostraremos como a programação geométrica pode ser utilizada no cálculo da raiz de Perron. Posteriormente faremos uma conexão daquilo que foi desenvolvido até o momento com o formalismo termodinâmico.

As principais referências utilizadas na elaboração deste capítulo, foram: [3], [10] e [11].

3.1 Estimativa da raiz de Perron através da programação geométrica

Seja $B = [b_{ij}] \in M_{n\times n}(\mathbb{R})$, uma matriz positiva, $B > 0$. De acordo com o teorema 1.2.8, existe $r = \rho(B)$ (raiz de Perron), tal que, $r > 0, B(p) = rp$ e $||p||_1 = 1$, de modo que $p > 0$.

Desta forma, pela segunda igualdade escrita na linha anterior, temos:

$$\rho(B) = \sum_{j=1}^{n} \frac{b_{ij}p_j}{p_i}, \quad i = 1, 2, ..., n. \tag{3.1}$$

Consideremos o seguinte problema de otimização:

PROGRAMA PRIMAL

$$\text{"} \min x_0 \tag{3.2}$$

sujeito as seguintes restrições:

$$x_0 \geq \sum_{j=1}^{n} \frac{b_{ij}x_j}{x_i}, \quad i = 1, 2, ..., n. \tag{3.3}$$

$$x_j > 0, \quad j = 0, 1, 2, ..., n.\text{"} \tag{3.4}$$

Afirmamos que a solução otimizada para tal problema é $\rho(B)$. De fato:

Seja x_0, x_1, ..., x_n qualquer solução possível para o problema (3.2) - (3.4). Pelo teorema (1.2.7), para qualquer $x > 0$, temos:

$$\min_{1 \leq i \leq n} \frac{1}{x_i} \sum_{j=1}^{n} b_{ij}x_j \leq \rho(B)$$

Agora, seja $p = (p_1, p_2, ..., p_n)$, o vetor de Perron de B. Podemos determinar escalares positivos, $\alpha_i, i \in \{1, 2, ..., n\}$, tais que, $\alpha_i p_i = x_i$. Seja, $\alpha_{i_0} = \min_{1 \leq i \leq n} \alpha_i$, de forma que i_0 denota o índice onde tal mínimo ocorre. Deste modo, temos:

$$\frac{\sum_{j=1}^{n} b_{i_0 j}x_j}{x_{i_0}} \geq \frac{\sum_{j=1}^{n} b_{i_0 j}\alpha_{i_0}p_j}{\alpha_{i_0}p_{i_0}} = \frac{\sum_{j=1}^{n} b_{i_0 j}p_j}{p_{i_0}} = \rho(B)$$

Disto, segue-se que:

$$\max_{1 \leq i \leq n} \frac{1}{x_i} \sum_{j=1}^{n} b_{ij}x_j \geq \rho(B). \tag{3.5}$$

Consequentemente, juntando as duas desigualdades acima, temos que, para qualquer $x > 0$, é válida a seguinte cadeia de desigualdades:

$$\min_{1 \leq i \leq n} \frac{1}{x_i} \sum_{j=1}^{n} b_{ij}x_j \leq \rho(B) \leq \max_{1 \leq i \leq n} \frac{1}{x_i} \sum_{j=1}^{n} b_{ij}x_j \tag{3.6}$$

Por (3.5), concluímos que:

$$\rho(B) \leq x_0. \tag{3.7}$$

Agora, seja $x_0^*, x_1^*, ..., x_n^*$ a solução otimizada do problema (3.2) - (3.4). De acordo com (3.1) $\rho(B)$ é uma solução para o problema de otimização citado acima, logo $\rho(B) \geq x_0^*$. Por outro lado, de (3.7), concluímos que, $\rho(B) \leq x_0^*$. Portanto, $\rho(B) = x_0^*$.

Note que o problema (3.2) - (3.4) pode ser reformulado da seguinte maneira:
PROGRAMA PRIMAL

$$\text{"} \min x_0, \tag{3.8}$$

sujeito as seguintes restrições:

$$1 \geq \sum_{j=1}^{n} \frac{b_{ij}x_j}{x_i x_0}, \quad i = 1, 2, ..., n. \tag{3.9}$$

$$x_j > 0, \quad j = 0, 1, 2, ..., n.\text{"} \tag{3.10}$$

Note que se trocarmos x_0 por α em (3.3) a restrição pode ser reescrita como

$$(B - \alpha I)x \leq 0 \leftrightarrow \sum_{j=1}^{n} b_{ij}x_j \leq \alpha x_i \leftrightarrow \sum_{j=1, j\neq i}^{n} b_{ij}x_j + (b_{ii} - \alpha)x_i \leq 0, \quad i = 1, 2, ..., n.$$

O seguinte programa será útil na estimativa do autovalor de Perron.

$$\text{"}\mathit{Determine} \max_{k\in\{1,2,...,n\}} x_k, \tag{3.11}$$

sujeito as restrições:

$$\sum_{j=1, j\neq i}^{n} b_{ij}x_j + (b_{ii} - \alpha)x_i \leq 0, \quad i = 1, 2, ..., n, \tag{3.12}$$

$$x_j \geq 0, \quad j = 1, 2, ..., n.\text{"} \tag{3.13}$$

Mostraremos que este programa tem ou solução nula ou solução infinita, dependendo do valor de α.

Lema 3.1.1. *Seja B uma matriz quadrada, $B = [b_{ij}] \in M_{n\times n}(\mathbb{R})$, $B > 0$, e α um número real não negativo. Nestas condições, $\rho(B) > \alpha$, se e somente se,*

$$\sum_{j=1, j\neq i}^{n} b_{ij}x_j + (b_{ii} - \alpha)x_i \leq 0, \quad i = 1, 2, ..., n. \tag{3.14}$$

tem como solução não negativa somente a solução nula.

Demonstração: ($\Rightarrow$) Suponhamos por absurdo que $x = (x_1, ..., x_n)$ seja uma solução não-negativa do sistema (3.14) com $x \neq 0$. Considere o conjunto $J = \{k \in \{1, 2, ..., n\} | \ x_k \neq 0\}$. Inicialmente afirmamos que J assim definido é uma parte própria de {1,2,...,n}. De fato, suponhamos por absurdo que exista $x > 0$, tal que, (3.14) seja satisfeita. Logo, existe $x > 0$, tal que, $\frac{\sum\limits_{j=1}^{n} b_{ij}x_j}{x_i} \leq \alpha < \rho(B)$, o que contradiz (3.5), portanto concluímos a afirmação previamente proferida.

Tomemos $i \notin J$. Por (3.14), temos, $\sum\limits_{j=1}^{n} b_{ij}x_j \leq \alpha x_i$ e como $i \notin J$, segue-se da definição do conjunto J, que $\sum\limits_{j=1}^{n} b_{ij}x_j \leq 0$, ou equivalentemente, $\sum\limits_{j\in J} b_{ij}x_j \leq 0, i \notin J$. Isso significa que $b_{ij} = 0$ para todo $i \notin J, j \in J$, o que contraria nossa hipótese. Portanto, $J = \emptyset$ e consequentemente, a única solução para (3.14) é a solução nula.
($\Leftarrow$) Consideremos $(B - \alpha I)x \leq 0$ de modo que a única solução não negativa é a solução nula. Suponhamos por absurdo que $\alpha \geq \rho(B)$. Ora, mas se isto ocorresse, o vetor de Perron seria uma solução para $(B - \alpha I)x \leq 0$ o que contradiz a hipótese de que a única solução existente é a solução nula. Portanto, $\rho(B) > \alpha$.

■

Lema 3.1.2. *Sejam $B = [b_{ij}] \in M_{n\times n}(\mathbb{R})$, $B > 0$, e α um número real não negativo. Nestas condições, $(B - \alpha I)x \leq 0$ tem como solução não negativa somente a solução nula, se e somente se, a solução do problema (3.11)- (3.13) é 0.*

Demonstração: ($\Rightarrow$) Evidente.

($\Leftarrow$) Consideremos o problema (3.11)- (3.13) de modo que sua solução é 0. Suponhamos por absurdo que $(B - \alpha I)x \leq 0$ tenha alguma solução não negativa diferente da solução nula. Seja x_0 tal solução. Ora, como x_0 diferente da solução trivial, existe uma coordenada x_p em x_0, tal que $x_p > 0$. Claramente isso contradiz a hipótese de a solução otimizada do problema (3.11)- (3.13) é 0. ■

Portanto pelos lemas (3.1.1) e (3.1.2), concluímos que, o programa (3.11)- (3.13) apresenta solução otimizada nula se e somente se $\alpha < \rho(B)$.

Seja $\alpha \geq \rho(B)$ e $p > 0$ o vetor de Perron associado ao autovalor $\rho(B)$, então temos:

$$\sum_{j=1}^{n} \frac{b_{ij}p_j}{p_i} = \rho(B) \leq \alpha, \quad i = 1, 2, ..., n.$$

Mais que isso, para qualquer $\gamma > 0$, o vetor γp é solução da desigualdade acima. Logo se $\alpha \geq \rho(B)$, então o programa (3.11)- (3.13) não tem um máximo finito, pois para qualquer vetor da forma γp, com $\gamma > 0$ satisfaz as restrições do problema, e podemos tomar γ suficientemente grande, de modo que todas as coordenadas do vetor γp tornem-se tão grande quanto se deseje.

Donde concluímos que o programa (3.11)- (3.13) ou tem solução nula ou infinita.

Baseado nas ideias desenvolvidas até o momento, podemos determinar o seguinte algoritmo de aproximações sucessivas para o cálculo de $\rho(B)$.

Fixemos:

$$\rho' = \min_{1 \leq i \leq n} \sum_{j=1}^{n} b_{ij}, \quad \rho" = \max_{1 \leq i \leq n} \sum_{j=1}^{n} b_{ij}$$

Tomemos $x = \sum_{i=1}^{n} e_i$, onde e_i, denota o i-ésimo vetor da base canônica de $\mathbb{R}^n$. Pela desigualdade (3.6), temos que,

$$\rho' \leq \rho(B) \leq \rho".$$

Note que, se $\rho' = \rho"$, então, $\rho(B) = \rho'$ e não temos nada mais a fazer.

Caso a igualdade não ocorra, procedemos da seguinte maneira visando calcular o autovalor $\rho(B)$ com uma precisão ε.

Definimos $\rho'^{(1)} = \rho'$ e $\rho"^{(1)} = \rho"$ e $\rho^{*(1)} = \frac{\rho'^{(1)}+\rho"^{(1)}}{2}$. Para definir $\rho'^{(2)}$ e $\rho"^{(2)}$ usaremos o programa (3.11)- (3.13) na tomada de decisão da seguinte maneira. Tomemos $\alpha = \rho^{*(1)}$ no problema (3.11)- (3.13). Se a solução otimizada é a solução nula, então, (pelos lemas provados acima) defina $\rho'^{(2)} = \rho^{*(1)}$, e $\rho"^{(2)} = \rho"^{1}$. Caso contrário, isto é, se $x_k \to \infty$ no problema (3.11)- (3.13), defina $\rho'^{(2)} = \rho'^{(1)}$, e $\rho"^{(2)} = \rho^{*(1)}$, da mesma forma $\rho^{*(2)} = \frac{\rho'^{(2)}+\rho"^{(2)}}{2}$ e assim prosseguimos com este método de bissecção. O processo termina quando $\rho"^{(m)} - \rho'^{(m)} < \varepsilon$. Assim conseguimos o cálculo de $\rho(B)$ com uma precisão ε. O valor aproximado de $\rho(B)$ é dado por:

$$\rho(B) \approx \frac{\rho'^{(m)} + \rho"^{(m)}}{2}.$$

Exemplo 3.1.3. *Dada* $A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 4 \end{bmatrix}$. *Estime a raiz de perron de A, com uma precisão* $\varepsilon = 0,1$.

De acordo com a notação anterior, $\rho^{'(1)} = 4$, e $\rho^{''(1)} = 10$. Assim, como $\rho^{*(1)} = \frac{\rho^{'(1)}+\rho^{''(1)}}{2}$, segue-se que, $\rho^{*(1)} = 7$. Tome $\alpha = \rho^{*(1)} = 7$, no problema (3.11)- (3.13). Através de um comando chamado linprog presente no Matlab é possível constatar que a solução otimizada do problema (3.11)- (3.13) é zero. Assim, de acordo com os lemas provados acima, temos que, $\rho^{*(1)} \leq \rho(A) \leq \rho^{''(1)}$.

Façamos, $\rho^{'(2)} = \rho^{*(1)}$, $\rho^{''(2)} = \rho^{''(1)}$ donde vem que, $\rho^{*(2)} = \frac{\rho^{'(2)}+\rho^{''(2)}}{2} = 8,5$. Tome $\alpha = \rho^{*(2)} = 8.5$, no problema (3.11)- (3.13). Através de um comando chamado linprog presente no programa Matlab é possível constatar que a solução otimizada do problema (3.11)- (3.13), $x_k \to \infty$. Assim, de acordo com os lemas provados acima, temos que, $\rho^{'(2)} \leq \rho(A) \leq \rho^{*(2)}$.

Façamos, $\rho^{'(3)} = \rho^{'(2)}$, $\rho^{''(3)} = \rho^{*(2)}$ donde vem que, $\rho^{*(3)} = \frac{\rho^{'(3)}+\rho^{''(3)}}{2} = 7,75$. Tome $\alpha = \rho^{*(3)} = 7,75$, no problema (3.11)- (3.13). Através de um comando chamado linprog presente no programa Matlab é possível constatar que a solução otimizada do problema (3.11)- (3.13), $x_k \to \infty$. Assim, de acordo com os lemas provados acima, temos que, $\rho^{'(3)} \leq \rho(A) \leq \rho^{*(3)}$.

Façamos, $\rho^{'(4)} = \rho^{'(3)}$, $\rho^{''(4)} = \rho^{*(3)}$ donde vem que, $\rho^{*(4)} = \frac{\rho^{'(4)}+\rho^{''(4)}}{2} = 7,375$. Tome $\alpha = \rho^{*(4)} = 7,375$, no problema (3.11)- (3.13). Através de um comando chamado linprog presente no programa Matlab é possível constatar que a solução otimizada do problema (3.11)- (3.13), 0. Assim, de acordo com os lemas provados acima, temos que, $\rho^{*(4)} \leq \rho(A) \leq \rho^{''(4)}$.

Façamos, $\rho^{'(5)} = \rho^{*(4)}$, $\rho^{''(5)} = \rho^{''(4)}$ donde vem que, $\rho^{*(5)} = \frac{\rho^{'(5)}+\rho^{''(5)}}{2} = 7,5625$. Tome $\alpha = \rho^{*(5)} = 7,5625$, no problema (3.11)- (3.13). Através de um comando chamado linprog presente no programa Matlab é possível constatar que a solução otimizada do problema (3.11)- (3.13), $x_k \to \infty$. Assim, de acordo com os lemas provados acima, temos que, $\rho^{'(5)} \leq \rho(A) \leq \rho^{*(5)}$.

Façamos, $\rho^{'(6)} = \rho^{'(5)}$, $\rho^{''(6)} = \rho^{*(5)}$ donde vem que, $\rho^{*(6)} = \frac{\rho^{'(6)}+\rho^{''(6)}}{2} = 7,46875$. Tome $\alpha = \rho^{*(6)} = 7,46875$, no problema (3.11)- (3.13). Através de um comando chamado linprog presente no programa Matlab é possível constatar que a solução otimizada do problema (3.11)- (3.13), $x_k \to \infty$. Assim, de acordo com os lemas provados acima, temos que, $\rho^{'(6)} \leq \rho(A) \leq \rho^{*(6)}$.

Façamos, $\rho^{'(7)} = \rho^{'(6)}$, $\rho^{''(7)} = \rho^{*(6)}$ donde vem que, $\rho^{*(7)} = \frac{\rho^{'(7)}+\rho^{''(7)}}{2} = 7,421875$. Tome $\alpha = \rho^{*(7)} = 7,421875$, no problema (3.11)- (3.13). Através de um comando chamado linprog presente no programa Matlab é possível constatar que a solução otimizada do problema (3.11)- (3.13), é 0. Assim, de acordo com os lemas provados acima, temos que, $\rho^{*(7)} \leq \rho(A) \leq \rho^{''(7)}$.

Façamos, $\rho^{'(8)} = \rho^{*(7)}$, $\rho^{''(8)} = \rho^{''(7)}$. O processo termina aqui, pois $\rho^{''(8)} - \rho^{'(8)} =$

$0,046875 < 0,1$. Assim, assumimos que:

$$\rho(B) \approx \frac{\rho'^{(8)} + \rho"^{(8)}}{2} = 7,4453125$$

O algoritmo descrito acima é enunciado a seguir.

Algoritmo.

Seja $B \in M_{n\times n}(\mathbb{R})$, tal que, $B > 0$.

Dados iniciais: $\rho' = \min\limits_{1\leq i\leq n} \sum\limits_{j=1}^{n} b_{ij}$, $\rho" = \max\limits_{1\leq i\leq n} \sum\limits_{j=1}^{n} b_{ij}$ e precisão $\varepsilon > 0$.

Passo 1 - k=1.

Passo 2 - Defina $\rho^{*(k)} = \frac{\rho'+\rho''}{2}$.

Passo 3 - Se $\rho'' - \rho^{*(k)} < \frac{\varepsilon}{2}$, então faça $\rho(A) = \rho^{*(k)}$. Fim.

Passo 4 - Se ao tomarmos $\alpha = \rho^{*(k)}$ no problema (3.11)- (3.13) e este tiver 0 como solução otimizada, faça $\rho' = \rho^{*(k)}$. Caso contrário, $\rho'' = \rho^{*(k)}$.

Passo 5 - Faça k=k+1.

Passo 6 - Volte ao passo 2.

3.2 Formulação do cálculo da raiz de Perron como um problema de programação geométrica

Seja $B = [b_{ij}] \in M_{n\times n}(\mathbb{R})$, uma matriz positiva, $B > 0$. Vimos na seção 3.1 que para calcular a raiz de Perron, basta resolver o

PROGRAMA PRIMAL:

$$\text{"} \min x_0,$$

sujeito as seguintes restrições:

$$1 \geq \sum_{j=1}^{n} \frac{b_{ij}x_j}{x_i x_0}, \quad i = 1, 2, ..., n.$$

$$x_j > 0, \quad j = 0, 1, 2, ..., n.\text{"}$$

A seguir reformulamos o programa primal acima na linguagem da programação geométrica conforme visto na Definição 2.1.7.

PROGRAMA PRIMAL:

$$\text{"} \min g_0(x) = x_0,$$

sujeito as seguintes restrições:

$$x_0 > 0,\ x_1 > 0,\ ..., x_n > 0.$$

e

$$g_1(x) \leq 1,\ g_2(x) \leq 1,\ ..., g_n(x) \leq 1$$

de modo que,

$$g_k(x) = \sum_{j=1}^{n} b_{kj} x_0^{-1} x_k^{-1} x_j,\ \ k = 1, ..., n.$$

Para colocarmos o problema na notação usada anteriormente definimos,

$$(c_1, c_2, c_3, ..., c_{n^2+1}) = (1, b_{11}, b_{12}, ..., b_{nn})$$

onde $c_1 = 1$, $b_{ij} = c_t$ para $t = j + (i-1)n + 1,\ \ i, j = 1, ..., n$ e $J[0] = \{1\}$, $J[1] = \{2, ..., n+1\}, J[2] = \{n+2, ..., 2n+1\}, ..., J[n] = \{n^2 - n + 2, ..., n^2 + 1\}$."

A cada $k = 0, 1, ..., n$ vamos associar uma matriz utilizando o procedimento abaixo.

Para $g_0(x)$ associamos a matriz E_0 de dimensão $1 \times (n+1)$ cujas componentes são os expoentes das variáveis $x_0, x_1, ..., x_n$, respectivamente, onde os índices estão ordenados de forma crescente. Desta forma:

$$E_0 = \begin{cases} 1, & j = 1 \\ 0, & j > 1 \end{cases} \tag{3.15}$$

Para o posinômio $g_1(x)$ associamos a matriz E_1 de dimensão $n \times (n+1)$, de modo que a i-ésima linha é dada em função da i-ésima parcela do posinômio, mais especificamente, é fornecida pelos expoentes das variáveis $x_0, x_1, ..., x_n$ presentes nesta parcela, onde os indices estão ordenados de forma crescente. A formalização do programa primal A relembrada acima nos permite inferir que E_1 é dada pela seguinte matriz:

$$E_1 = \begin{cases} -1, & j = 1,\ i \in \{1, ..., n\} \\ -1, & j = 2,\ i \in \{1, ..., n\} - \{1\} \\ 1, & j = i + 1;\ i \neq 1 \\ 0, & demais\ casos \end{cases} \tag{3.16}$$

A título de ilustração, por exemplo se estivéssemos trabalhando com x_0, x_1, x_2, x_3 Nossas matrizes E_0 e E_1 seriam dadas por:

$$E_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}$$

e

$$E_1 = \begin{bmatrix} -1 & 0 & 0 & 0 \\ -1 & -1 & 1 & 0 \\ -1 & -1 & 0 & 1 \end{bmatrix}$$

Generalizando: Como temos $n+1$ posinômios, teremos $n+1$ matrizes. Para cada s, $1 \leq s \leq n$, construímos a matriz E_s de maneira análoga a que empregamos na concepção da matriz E_1. A ordem de E_s associada ao posinômio $g_s(x)$ será

$n \times (n+1)$. Sua lei de formação é:

$$E_s = \begin{cases} -1, & j = 1,\ i \in \{1, ..., n\} \\ -1, & j = s+1,\ i \in \{1, ..., n\} - \{s\} \\ 1, & j = i+1;\ i \neq s \\ 0, & demais\ casos \end{cases} \tag{3.17}$$

A matriz de expoentes do programa primal A, é dada na representação em blocos por:

$$E = \begin{bmatrix} E_0 \\ \cdots\cdots \\ E_1 \\ \cdots\cdots \\ \vdots \\ \cdots\cdots \\ E_n \end{bmatrix}. \tag{3.18}$$

Denotaremos a matriz acima por $E = [e_{ij}]_{(n^2+1)\times(n+1)}$.

Relacionado ao programa primal, temos o seguinte

PROGRAMA DUAL:

$$\text{"} \max v(\delta) = \max \left[\prod_{i=1}^{n^2+1} \left(\frac{c_i}{\delta_i} \right)^{\delta_i} \right] \prod_{k=1}^{n} \lambda_k(\delta)^{\lambda_k(\delta)}, \tag{3.19}$$

onde

$$\lambda_k(\delta) = \sum_{i \in J[k]} \delta_i,\ k = 1, ..., n,$$

$$(c_1, c_2, ..., c_{n^2+1}) = (1, b_{11}, ..., b_{nn}) \tag{3.20}$$

$c_1 = 1$, $b_{ij} = c_t$ para $t = j + (i-1)n + 1,\ i, j = 1, ..., n$ e $J[0] = \{1\}$, $J[1] = \{2, ..., n+1\}, J[2] = \{n+2, ..., 2n+1\}, ..., J[n] = \{n^2 - n + 2, ..., n^2 + 1\}$. E o vetor variável $\delta = (\delta_1, \delta_2, ..., \delta_{n^2+1})$ obedece as seguintes restrições:

$$\delta_1 \geq 0,\ \delta_2 \geq 0,\ ..., \delta_{n^2+1} \geq 0 \tag{3.21}$$

$$\sum_{i \in J[0]} \delta_i = 1 \Rightarrow \delta_1 = 1 \tag{3.22}$$

e

$$\sum_{i=1}^{n^2+1} e_{ij}\delta_i = 0, \quad j = 1, 2, ..., n+1,\text{”} \tag{3.23}$$

onde e_{ij} são as entradas da matriz de coeficientes E dada acima.

Note que a implicação apresentada em (3.22) é decorrente de $J[0] = 1$.

Claramente o vetor $(\rho(B)+1, p_1, ..., p_n)$ mostra que o programa primal A descrito acima é superconsistente. Além disso, como o seu valor mínimo é atingido em $(\rho(B), p_1, ..., p_n)$, podemos utilizar o teorema 2.1.13. Assim, para determinar a raiz de Perron, basta resolver o programa dual descrito acima. Para nossos propósitos vamos inicialmente reformular o programa dual B.

Definimos:

$$1 = \delta_1; \;\; z_{ij} = \delta_t \;\; para \;\; t = j + (i-1)n + 1, \quad i, j = 1, ..., n. \tag{3.24}$$

Disto segue-se que, $\lambda_k = \sum\limits_{j=1}^{n} z_{kj}, \;\; k = 1, ..., n$. Consequentemente, valendo-se destas últimas igualdades e aplicando (3.20) e (3.24) em (3.19) obtemos:

$$v(\delta_1, ..., \delta_{n^2+1}) = v(z_{ij}) = \left(\prod_{i=1}^{n}\left(\prod_{j=1}^{n}\left(\frac{b_{ij}}{z_{ij}}\right)^{z_{ij}}\right)\right)\left(\prod_{k=1}^{n}\left(\left(\sum_{j=1}^{n} z_{kj}\right)^{\sum\limits_{j=1}^{n} z_{kj}}\right)\right)$$

$$= \prod_{i=1}^{n}\prod_{j=1}^{n}\left(\frac{b_{ij}}{z_{ij}}\right)^{z_{ij}} \prod_{k=1}^{n}\left(\sum_{j=1}^{n} z_{kj}\right)^{\sum\limits_{j=1}^{n} z_{kj}}.$$

Observamos que a condição de ortogonalidade dada em (3.23) pode ser expressa pela seguinte igualdade de matrizes:

$$\delta^T E = 0^T, \quad \delta \in \mathbb{R}^{n^2+1}, \, 0 \in \mathbb{R}^{n+1}. \tag{3.25}$$

Observando a primeira sentença de (3.15) e de (3.17), bem como a primeira coluna de (3.18) aliada as relações (3.24) e (3.25), temos:

$$1 - \sum_{i=1}^{n}\sum_{j=1}^{n} z_{ij} = 0, \tag{3.26}$$

Agora, fixemos l, $1 < l \leq n+1$. Note que a l-ésima coluna de E é dada por

$$[E]_l = \begin{bmatrix} [E_0]_l \\ \cdots\cdots\cdots \\ [E_1]_l \\ \cdots\cdots\cdots \\ \vdots \\ \cdots\cdots\cdots \\ [E_n]_l \end{bmatrix},$$

onde $[E_u]_l$ denota a l-ésima coluna de E_u, $u \in \{0, 1, ..., n\}$. Por (3.15), segue-se que, $[E_0]_l = 0$. De acordo com (3.17),

$$[E_{l-1}]_l = \begin{cases} -1; & i \in \{1, ..., n\} - \{l-1\} \\ 0; & i = l-1 \end{cases} \tag{3.27}$$

Para $k \neq 0$ e $k \neq l-1$, (3.17), nos mostra que

$$[E_k]_l = \begin{cases} 1; & i = l-1 \\ 0; & i = \{1, ..., n\} - \{l-1\} \end{cases} \tag{3.28}$$

De acordo com (3.25), temos que $\delta^T [E]_l = 0$. Assim, usando (3.28), (3.27) e (3.24) temos:

$$- \sum_{j=1; j\neq(l-1)}^{n} z_{(l-1)j} + \sum_{j=1; j\neq(l-1)}^{n} z_{j(l-1)} = 0.$$

Como a equação acima é válida para qualquer l que obedeça a restrição, $1 < l \leq n+1$, temos:

$$- \sum_{j=1; j\neq i}^{n} z_{ij} + \sum_{j=1; j\neq i}^{n} z_{ji} = 0, \qquad i = 1, 2, ..., n. \tag{3.29}$$

De (3.29), vem que,

$$\sum_{j=1; j\neq i}^{n} z_{ji} = \sum_{j=1; j\neq i}^{n} z_{ij}, \qquad i = 1, 2, ..., n.$$

Da condição de positividade (3.21), e usando a relação (3.24) temos que,

$$z_{ij} \geq 0, \qquad i, j = 1, 2, ..., n.$$

Portanto, resolver o programa dual correspondente ao programa primal dado, equivale a resolver o problema abaixo:

PROGRAMA DUAL:

$$\text{”}\max v(z_{ij}) = \max \left(\prod_{i=1}^{n} \left(\prod_{j=1}^{n} \left(\frac{b_{ij}}{z_{ij}} \right)^{z_{ij}} \right) \right) \left(\prod_{k=1}^{n} \left(\left(\sum_{j=1}^{n} z_{kj} \right)^{\sum_{j=1}^{n} z_{kj}} \right) \right) \tag{3.30}$$

sujeito as restrições:

$$z_{ji} \geq 0, \qquad i, j = 1, 2, ..., n. \tag{3.31}$$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} z_{ij} = 1. \tag{3.32}$$

$$\sum_{j=1;j\neq i}^{n} z_{ji} = \sum_{j=1;j\neq i}^{n} z_{ij}, \qquad i = 1, 2, ..., n.\text{”} \tag{3.33}$$

Agora pelo lema 2.1.12 que os problemas primal e dual atingem o mesmo valor, isto é, $v(\delta^*) = g_0(x^*)$, se e somente se,

$$\delta^*{}_t = \begin{cases} \frac{c_t {x^*}_1^{a_{t1}} {x^*}_2^{a_{t2}} ... {x^*}_n^{a_{tn}}}{g_0(x^*)}, & t \in J[0] \\ \lambda_k(\delta^*) c_t {x^*}_1^{a_{t1}} {x^*}_2^{a_{t2}} ... {x^*}_n^{a_{tn}}, & t \in J[k]; \quad k = 1, 2, ..., n \end{cases}$$

De acordo com (3.24), tomando $t = j + (i-1)n + 1$, a equivalência acima torna-se:

$$v(z^*{}_{ij}) = g_0(x^*) \Leftrightarrow \begin{bmatrix} 1 \\ z^*{}_{ij} \end{bmatrix} = \begin{bmatrix} 1 \\ \left(\sum_{v=1}^{n} z^*{}_{iv} \right) b_{ij} {x^*}_0^{-1} {x^*}_i^{-1} x^*{}_j \end{bmatrix}.$$

3.3 A programação geométrica vista como uma ferramenta para o formalismo termodinâmico

Nesta seção usamos o programa dual B associado a resultados bem conhecidos em formalismo termodinâmico para exibir a medida de Gibbs associada a um potencial que depende apenas das duas primeiras coordenadas.

Seja $B = [b_{ij}] \in M_{n\times n}(\mathbb{R})$, uma matriz positiva, $B > 0$. Vimos na seção 3.1 que para calcular a raiz de Perron, basta resolver o

PROGRAMA PRIMAL:

$$\text{"} \min x_0,$$

sujeito as seguintes restrições:

$$1 \geq \sum_{j=1}^{n} \frac{b_{ij}x_j}{x_i x_0}, \qquad i = 1, 2, ..., n.$$

$$x_j > 0, \qquad j = 0, 1, 2, ..., n.\text{"}$$

Também constamos, na seção 3.2, que a este programa primal, temos um programa Dual relacionado, e que solucionar o programa dual, consiste em resolver o problema dado abaixo:

PROGRAMA DUAL:

$$\text{"} \max v(z_{ij}) = \max \left(\prod_{i=1}^{n} \left(\prod_{j=1}^{n} \left(\frac{b_{ij}}{z_{ij}} \right)^{z_{ij}} \right) \right) \left(\prod_{k=1}^{n} \left(\left(\sum_{j=1}^{n} z_{kj} \right)^{\sum\limits_{j=1}^{n} z_{kj}} \right) \right), \tag{3.34}$$

sujeito as restrições:

$$z_{ji} \geq 0, \qquad i, j = 1, 2, ..., n. \tag{3.35}$$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} z_{ij} = 1. \tag{3.36}$$

$$\sum_{j=1}^{n} z_{ji} = \sum_{j=1}^{n} z_{ij}, \qquad i = 1, 2, ..., n.\text{"} \tag{3.37}$$

Note que

$$\prod_{k=1}^{n} \left(\sum_{l=1}^{n} z_{kl} \right)^{\sum_{l=1}^{n} z_{kl}} = \prod_{k=1}^{n} \prod_{j=1}^{n} \left(\sum_{l=1}^{n} z_{kl} \right)^{z_{kj}} = \prod_{i=1}^{n} \prod_{j=1}^{n} \left(\sum_{l=1}^{n} z_{il} \right)^{z_{ij}},$$

logo a equação (3.34) pode ser reescrita da seguinte forma:

$$v(z_{ij}; i,j=1,..,n) = \prod_{i=1}^{n}\prod_{j=1}^{n}\left(\frac{b_{ij}\ \sum_{l=1}^{n} z_{il}}{z_{ij}}\right)^{z_{ij}} . \tag{3.38}$$

Queremos interpretar as restrições (3.35) a (3.37) como uma cadeia de Markov. Com este intuito definimos $\pi_j := \sum_{i=1}^{n} z_{ij} \geq 0$.

Pela equação (3.37) temos que $\pi_j = \sum_{i=1}^{n} z_{ij} = \sum_{i=1}^{n} z_{ji}$, portanto trocando os nomes das variáveis temos que $\pi_i = \sum_{j=1}^{n} z_{ij}$. Disto segue que $\pi_i = 0$ se e somente se $z_{ij} = 0$ para todo $j \in \{1,...,n\}$. Agora podemos definir

$$P_{ij} := \begin{cases} \frac{z_{ij}}{\pi_i}, & se \quad \pi_i > 0, \\ P_{ij};\ tal\,que,\ P_{ij} \geq 0,\ \sum_{j=1}^{n} P_{ij} = 1 & se \quad \pi_i = 0. \end{cases}$$

Assim $z_{ij} = \pi_i P_{ij}$. Portanto, quando $\pi_i > 0$, $\pi_i = \sum_{j=1}^{n} z_{ij} = \pi_i \sum_{j=1}^{n} P_{ij}$ o que implica em

$$\sum_{j=1}^{n} P_{ij} = 1.$$

E por definição, $\sum_{j=1}^{n} P_{ij} = 1$, quando $\pi_i = 0$. Assim, $\sum_{j=1}^{n} P_{ij} = 1$, em qualquer caso.

Por definição temos que

$$\pi_j := \sum_{i=1}^{n} \pi_i P_{ij}. \tag{3.39}$$

A equação (3.36) torna-se

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \pi_i P_{ij} = 1.$$

E finalmente a equação (3.35) implica em

$$\pi_i, P_{ij} \geq 0 .$$

Na nova notação o problema consiste em maximizar

$$v(\pi_i, P_{ij}; i,j=1,..,n) = \prod_{i=1}^{n}\prod_{j=1}^{n}\left(\frac{b_{ij}}{P_{ij}}\right)^{\pi_i P_{ij}} , \tag{3.40}$$

com as seguintes restrições

$$\sum_{j=1}^{n} P_{ij} = 1 \;\; ; \;\; \sum_{i=1}^{n}\sum_{j=1}^{n} \pi_i P_{ij} = 1 \;, \tag{3.41}$$

$$\sum_{i=1}^{n} \pi_i P_{ij} = \pi_j \;, \tag{3.42}$$

$$\pi_i, P_{ij} \geq 0 \;. \tag{3.43}$$

Sabemos, pelo teorema de Perron, que o programa Primal tem solução estritamente positiva $(x_0^*, x_1^*, ..., x_n^*)$, portanto pelo teorema 2.1.13 temos que o problema (3.19) - (3.23) tem solução otimal. Donde vem que, o problema (3.30) - (3.33) também é solúvel, e consequentemente, o problema (3.40) - (3.43) também é solúvel.

Agora pelo lema 2.1.12 que os programas primal e dual atingem o mesmo valor, isto é, $v(\delta^*) = g_0(x^*)$, se e somente se,

$$\delta^*{}_t = \begin{cases} \frac{c_t x^*{}_1^{a_{t1}} x^*{}_2^{a_{t2}} ... x^*{}_n^{a_{tn}}}{g_0(x^*)}, & t \in J[0] \\ \lambda_k(\delta^*) c_t x^*{}_1^{a_{t1}} x^*{}_2^{a_{t2}} ... x^*{}_n^{a_{tn}}, & t \in J[k]; \quad k = 1, 2, ..., n \end{cases}$$

De acordo com (3.24), tomando $t = j + (i-1)n + 1$, a equivalência acima torna-se:

$v(z^*{}_{ij}) = g_0(x^*)$, se e somente se,

$$\begin{bmatrix} 1 \\ z^*{}_{ij} \end{bmatrix} = \begin{bmatrix} 1 \\ \left(\sum_{v=1}^{n} z^*{}_{iv}\right) b_{ij} x^*{}_0^{-1} x^*{}_i^{-1} x^*{}_j \end{bmatrix}.$$

Passando ao problema (3.40) - (3.43), devemos determinar (π^*, P^*), de modo que, $v(\pi^*, P^*) = g_0(x^*)$. Sendo $z^*{}_{ij} = \pi^*{}_i P^*{}_{ij}$temos que a igualdade ocorre, se e somente se,

$$P_{ij}^* = \frac{b_{ij} \;\; x_j^*}{x_0^* x_i^*}.$$

Vamos interpretar os resultados acima usando formalismo termodinâmico para cadeias de Markov.

Dizemos que uma matriz $P \in M_{n\times n}$ com entradas não-negativas P_{ij} é estocástica se $\sum_{j=1}^{n} P_{ij} = 1$ (restrição (3.41)). Uma matriz estocástica descreve uma cadeia de Markov com n estados, onde P_{ij} representa a probabilidade de passagem do estado i para o estado j. Chamamos um vetor π de vetor de probabilidade estacionário se π é um autovetor à esquerda da matriz P associado ao autovalor 1, i.e. $\pi P = \pi$ (restrição (3.42)).

Considere o espaço de Bernoulli dado pelas sequências $\Sigma := \{1, ..., n\}^{\mathbb{N}}$, ou seja, $x \in \Sigma$ então $x = (x_1, x_2, ...)$, com $x_i \in \{1, ..., n\}$. Seja $0 < \theta < 1$, definimos uma métrica d em Σ dada por:

$$d(x,y) = \begin{cases} 1, & se\, x_1 \neq y_1 \\ \theta^N, & N = \max\{j \in \mathbb{N} | 1 \leq i < j \;\; e \;\; x_i = y_i\} \end{cases}$$

Tomemos agora em Σ a dinâmica dada pelo shift, $\sigma : \Sigma \to \Sigma$ definido por $\sigma(x_1, x_2, ...) = (x_2, x_3, ...)$. Seja $\mathcal{M}_\sigma$ as medidas invariantes pelo shift, i.e., $\mu \in \mathcal{M}_\sigma$ se $\mu(B) = \mu(\sigma^{-1}(B))$, para todo B Boreliano de Σ.

Os conjuntos

$$[i_1...i_k] = \{x \in \Sigma : x_1 = i_1, ..., x_k = i_k\}$$

são chamados de cilindros em Σ.

À cadeia de Markov, representada por P e π, associamos a medida de Markov em Σ que, nos cilindros, é dada por

$$\mu([i_1...i_k]) = \pi_{i_1} P_{i_1 i_2} P_{i_2 i_3} ... P_{i_{k-1} i_k}.$$

De acordo com a seção 7.3 de [11], temos que a medida μ assim definida é σ-invariante.

Seja μ uma medida em Σ, pela seção 8.3 de [11] podemos definir a entropia da medida μ como

$$h(\mu) = \lim_{k\to\infty} -\frac{1}{k} \sum_{[i_1...i_k]} \mu([i_1...i_k]) \log \mu([i_1...i_k]).$$

Sendo μ a medida de Markov representada por P e π, temos ainda pela seção 8.3 de [11] que

$$h(\mu) = -\sum_{i=1}^{n}\sum_{j=1}^{n} \pi_i P_{ij} \log P_{ij}.$$

Seja $f : \Sigma \to \mathbb{R}$ uma função que chamaremos de observável que descreve alguma grandeza física, por exemplo energia ou temperatura.

Em formalismo termodinâmico estamos interessados em encontrar uma medida que satisfaça o seguinte princípio variacional

$$P(f) = \sup_{\mu\in\mathcal{M}_\sigma} \left\{ \int f d\mu + h_\mu(\sigma) \right\}, \tag{3.44}$$

onde $P(f)$ é chamada de pressão de f. A medida que atinge este supremo é chamada de medida de Gibbs.

No caso em que f depende apenas das duas primeiras coordenadas do espaço Σ, dizemos que o potencial descreve a iteração de vizinhos, pois Σ pode ser pensado como um reticulado unidimensional $\mathbb{N}$ onde a cada natural está associado a um estado $1, ..., n$. Assim $f(x_1, x_2)$ leva em consideração apenas os valores nas posições vizinhas 1 e 2. Podemos associar a f uma matriz B $n \times n$ com entradas $b_{ij} = e^{f(i,j)}$. É conhecido (veja capítulos 2 e 3 de [10]) que neste caso $P(f) = \log \lambda_B$, onde λ_B é o maior autovalor da matriz B.

Voltando ao programa dual, temos que extraindo o logaritmo de $v(\pi, P)$, o programa dual equivale a maximizarmos, entre todas as medidas de Markov, a função

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\log\left(b_{ij}\right)\pi_i P_{ij} - \sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i P_{ij}\log\left(P_{ij}\right).$$

Seja agora o potencial $f : \Sigma \to \mathbb{R}$ dado por $f(x_1, x_2, x_3, ...) = f(x_1, x_2) = \log\left(b_{x_1 x_2}\right)$, que depende apenas das duas primeiras coordenadas de x.

Note que como f depende apenas das duas primeiras coordenadas de x

$$\int f d\mu = \sum_{i=1}^{n}\sum_{j=1}^{n} f(i,j)\ \pi_i P_{ij}.$$

Ou seja, o programa dual consiste em encontrar a medida de Markov que maximiza a energia livre do potencial f dado acima.

Portanto seja λ_B a raiz de Perron da matriz B temos que

$$\log\lambda_B = \max\left\{\int f d\mu + h_\mu(\sigma)\ \ |\ \ \mu \text{ é medida de Markov}\right\}.$$

Ainda sabemos que o máximo é atingido quando μ^* é determinada por

$$P_{ij}^* = \frac{b_{ij}\ x_j^*}{\lambda_B\, x_i^*}\quad e\quad \pi^*,$$

onde $(x_1^*, ..., x_n^*)$ é o autovetor positivo associado ao autovalor λ_B e π^* é o autovetor à esquerda da matriz P^*.

Assim

$$\log\lambda_B = \sum_{i=1}^{n}\sum_{j=1}^{n} f(i,j)\ \pi_i^* P_{ij}^* - \sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i^* P_{ij}^*\log\left(P_{ij}^*\right).$$

Portanto a medida de Markov μ^* definida através de P^* e π^* atinge o supremo em (3.44). Logo μ^* é a medida de Gibbs para o potencial f.

Apêndice A

A.1 O Lema de Farkas

O lema de Farkas aborda sistema de inequações lineares e é utilizado no desenvolvimento da teoria referente a programação geométrica. Entretanto, antes de demonstrá-lo, iremos apresentar alguns resultados auxiliares.

Definição A.1.1. *Seja E um espaço vetorial. Dados $a, b \in E$, o segmento de reta de extremos a e b é o conjunto $[a, b] = \{(1-t)a + tb; 0 \leq t \leq 1\}$.*

Definição A.1.2. *Um subconjunto X de um espaço vetorial E chama-se convexo quando*

$$a, b \in X \Rightarrow [a, b] \subset X.$$

Definição A.1.3. *Um subconjunto não vazio de $\mathbb{R}^n$, K, chama-se cone se a seguinte implicação é satisfeita.*

$$\forall \lambda > 0, \forall x \in K \Rightarrow \lambda x \in K$$

Definição A.1.4. *Seja X um subconjunto de $\mathbb{R}^n$. Um ponto $a \in X$ chama-se* um ponto interior *a X, quando existe $\delta > 0$ tal que, $||x - a|| < \delta \Rightarrow x \in X$.*

Definição A.1.5. *Um conjunto $X \subseteq \mathbb{R}^n$ chama-se aberto quando todos os seus pontos são interiores.*

Definição A.1.6. *Uma sequência em* $\mathbb{R}^n$ *é uma aplicação* $z : \mathbb{N} \to \mathbb{R}^n$*, definida no conjunto* $\mathbb{N}$ *dos naturais. O valor que essa aplicação assume no número* k *é indicado com* z_k *e chama-se o* k-ésimo termo *da sequência. Usaremos a notação* (z_k) *para indicar a sequência cujo k-ésimo termo é* $z_k \in \mathbb{R}^n$*.*

Definição A.1.7. *Dizemos que o ponto* $a \in \mathbb{R}^n$ *é limite da sequência de pontos* $z_k \in \mathbb{R}^n$ *quando para todo* $\varepsilon > 0$ *dado, é possível obter* $k_0 \in \mathbb{N}$ *tal que,*
$k > k_0 \Rightarrow ||z_k - a|| < \varepsilon$*. Neste caso, dizemos que* (z_k) *converge para* a *e escrevemos,* $\lim\limits_{k \to \infty} z_k = a$ *.*

Definição A.1.8. *Um ponto* $a \in \mathbb{R}^n$ *diz-se aderente a um conjunto* $X \subseteq \mathbb{R}^n$ *quando é limite de uma sequência de pontos desse conjunto.*

Definição A.1.9. *Um conjunto* $X \subseteq \mathbb{R}^n$ *chama-se fechado quando contém todos seus pontos aderentes.*

Teorema A.1.10. *Se* $A = (a_1\, a_2 \ldots a_n) \in M_{m \times n}(\mathbb{R})$*, onde* $a_1, \ldots, a_n$ *são as colunas de* A*, então,* $K = \{Ax\,;\, x \geq 0\}$ *é um cone convexo fechado.*

Demonstração: Seja $y_0 \in K$, e $\lambda > 0$. Pela definição de K, existe $x_0 \geq 0$, tal que, $y_0 = Ax_0$. Como:

$$\lambda y_0 = \lambda A x_0 = A(\lambda x_0)$$

e claramente, $\lambda x_0 \geq 0$, segue-se que, K é um cone.

Agora sejam $y_0, w_0 \in K, 0 \leq t \leq 1$. Pela definição de K, existem $x_0 \geq 0, v_0 \geq 0$, tal que, $y_0 = Ax_0$, $w_0 = Av_0$. Como:

$$ty_0 + (1-t)w_0 = tA(x_0) + (1-t)A(v_0) = A(tx_0) + A((1-t)w_0) = A(tx_0 + (1-t)w_0)$$

e evidentemente, $tx_0 + (1-t)w_0 \geq 0$, segue-se que, K é convexo.

Quando A é a matriz nula, temos que $K = \{0\}$ e o resultado é trivialmente verdadeiro.

Suponhamos que $A \neq 0$ e consideremos uma sequência, $\{y^k\}_{k\in\mathbb{N}}$ de elementos de K (ou seja, $y^k = A(x^k)$, com $x^k \geq 0$) de modo que tal sequência converge para y', devemos provar que $y' \in K$. Tal demonstração será dividida em dois casos:

Caso 1: Se as colunas de A, $a_1, ..., a_n$ são linearmente independentes, então para cada y^k existe um único x^k tal que $A(x^k) = y^k$. Sabemos de [4] que $A^T A$ é invertível. Deste modo, consideremos a aplicação continua $g : K \to \mathbb{R}^n_+$ definida por: $g(y) = (A^T A)^{-1} A^T(y)$. De [8], lembramos que a pré-imagem de um conjunto fechado por uma aplicação continua é um conjunto fechado. Assim, sendo
$\mathbb{R}^n_+ = \{x \in \mathbb{R}^n; x \geq 0\}$ um conjunto fechado, g continua, segue-se que, K é fechado, ou seja, $y' \in K$. Logo, existe $x' \in \mathbb{R}^n_+$, tal que, $A(x') = y'$.

Caso 2: Seja $v \in K$ e suponhamos que $v = A(u)$ (onde u e tal que as suas componentes são os escalares não negativos $\beta_1, \beta_2, ..., \beta_n$) e ressaltamos o conjunto de índices, $I_u = \{j : \beta_j > 0\}$, para os quais as colunas de A associadas a tal conjunto são linearmente dependentes. Assim existem escalares, não todos nulos (α_j com $j \in I_u$), tais que $\sum\limits_{j\in I_u} \alpha_j a_j = 0$, onde pelo menos um dos escalares α_j é positivo (pois caso contrário multiplica-se a equação anterior por -1). Definindo-se $\frac{\beta_p}{\alpha_p} = \min\left\{\frac{\beta_j}{\alpha_j};\ j \in I_u,\ \alpha_j > 0\right\}$ segue-se que:

$$v = \sum_{j\in I_u} \beta_j a_j - \frac{\beta_p}{\alpha_p}\sum_{j\in I_u} \alpha_j a_j = \sum_{j\in I_u}\left(\beta_j a_j - \frac{\beta_p}{\alpha_p}\alpha_j a_j\right)$$

$$= \sum_{j\in I_u - \{p\}}\left(\beta_j - \tfrac{\beta_p}{\alpha_p}\alpha_j\right) a_j \text{ com } \beta_j - \tfrac{\beta_p}{\alpha_p}\alpha_j \geq 0, \forall j \in I_u,$$

ou seja, $v \in K(I_u - p)$ (de modo que $K(S)$ denota o cone gerado pelas colunas de A cujos índices estão no conjunto S). Caso estas colunas sejam linearmente independentes paramos o processo. Caso contrário, repetimos o processo até obtermos uma conjunto de índices I_r, tal que o conjunto, $\{a_j; j \in I_r\}$ é composto apenas por vetores linearmente independentes e $v \in K(I_r)$. Pelo caso 1, $K(I_r)$ é um conjunto fechado. Denotemos por K_{11}, o cone convexo fechado formado pela primeira coluna não nula, K_{12} o cone convexo fechado formado pela segunda coluna não nula, e assim sucessivamente, teremos s_1 cones dessa forma, onde K_{1s_1} é o cone formado pela s_1-ésima coluna não nula. De modo genérico, dadas p colunas linearmente independentes, teremos s_p cones de modo que K_{pl_p}, $s_p \geq l_p \geq 1$, denota o l_p-ésimo cone gerado por p colunas linearmente independentes.

Denotemos por c o número de colunas linearmente independentes de A. Fazendo v variar em K, podemos obter K como a união dos cones convexos fechados $K_{11}, K_{12}, ..., K_{1s_1}, K_{21}, ..., K_{2s_2}, ..., K_{cs_c}$. Logo, $K = \bigcup_{j=1}^{c} \left(\bigcup_{i=1}^{s_j} K_{ji} \right)$. Então segue-se que K é fechado.

■

Teorema A.1.11. *Seja S um subconjunto fechado, convexo, não vazio de $\mathbb{R}^n$ e $x \in \mathbb{R}^n - S$. Então existe $c \in \mathbb{R}^n - \{0\}$ e existe $\varepsilon > 0$ tais que, $c^T x \geq c^T y + \varepsilon$, $\forall y \in S$.*

Demonstração: Seja S um subconjunto fechado, convexo, não vazio de $\mathbb{R}^n$ e $x \in \mathbb{R}^n - S$. Afirmamos que existe um ponto $y' \in S$ o qual é o mais próximo de x. De fato, uma vez que S é não vazio existe $z \in S$, agora consideremos a bola fechada de centro em x e raio $||x - z||$. Tal bola é um conjunto compacto. Ao fazermos intersecção desta bola com o conjunto fechado S obtemos um conjunto compacto, o qual denotaremos por A. Observando que a determinação do ponto de S que esta a distância mínima de x corresponde a descobrir o mínimo da função contínua $f(y) = ||x - y||$ no conjunto A. Pelo corolário de Weierstrass [7] vem que o ponto de mínimo, $y' \in S$, existe. Notamos ainda que $||x - y'|| > 0$, visto que, $y' \in S$ e $x \notin S$.

Agora, vamos mostrar que $2(x - y')^T (y' - y) \geq 0$, $\forall y \in S$. Ora, sendo y um ponto arbitrário do convexo S temos que $\forall\lambda \in (0,1)\,,\, \lambda y + (1-\lambda)y' \in S$. Relembrando que $y' \in S$, é o ponto de S, o qual está mais próximo de x, temos assim a seguinte desigualdade:

$$\|x-y'\|^2 \leq \|x-\lambda y-(1-\lambda)y'\|^2 = \|x-y'+\lambda(y'-y)\|^2$$

$$= \|x-y'\|^2 + 2\lambda(x-y')^T(y'-y) + \lambda^2\|y-y'\|^2$$

Assim:

$$2\left(x-y'\right)^T\left(y'-y\right)+\lambda\left\|y'-y\right\|^2 \geq 0 \Rightarrow \lim_{\lambda\to 0} 2(x-y')^T(y'-y)+\lambda\left\|y'-y\right\|^2 \geq 0$$

Ou seja:

$$2(x-y')^T(y'-y) \geq 0$$

Da desigualdade acima, segue-se que:

$$(x-y')^T(y'-y) \geq 0 \Leftrightarrow (x-y')^T(x-x+y'-y) = (x-y')^T(y'-x)+(x-y')^T(x-y) \geq 0.$$

Definindo: $c = x - y'$ e $\varepsilon = ||x-y'||^2$, obtemos o resultado desejado, conforme apresentamos abaixo.

$$-\varepsilon + c^T x - c^T y \geq 0 \Leftrightarrow c^T x \geq c^T y + \varepsilon$$

$\blacksquare$

Lema A.1.12. ***(Lema de Farkas)*** *Dada a matriz $A = (a_1\, a_2 \ldots a_n) \in M_{m\times n}(\mathbb{R})$ e $b \in \mathbb{R}^m$, então um e somente um dos seguintes sistemas associados tem solução.*

$$Ax = b \;\; e \;\; x \geq 0 \tag{A.1}$$

$$y^T A \leq 0 \;\; e \;\; y^T b > 0 \tag{A.2}$$

Demonstração: Suponhamos que (A.1) seja solúvel, isto é, existe $x' \in \mathbb{R}^n$ tal que, $Ax' = b$ com $x' \geq 0$. Assim, $y^T Ax' = y^T b$, $\forall y \in \mathbb{R}^m$. Deste modo, se existisse $y' \in \mathbb{R}^m$, tal que, $0 \geq y'^T A$, pela consequência (ii) da Definição 1.1.1 segue-se que $0 \geq y'^T Ax' = y'^T b$, ou seja, as desigualdades presentes em (A.2) não podem ser satisfeitas simultaneamente.

Agora, suponhamos que (A.1) não seja solúvel. Deste modo, infere-se que $b \notin K = \{Ax; x \geq 0\}$. De acordo com o teorema A.1.10 temos que K é um cone convexo fechado, e pelo teorema A.1.11, existem $y \in \mathbb{R}^m - \{0\}$ e $\varepsilon > 0$ tais que, $y^T b \geq y^T z + \varepsilon$, $\forall z \in K$. Observando-se que $0 \in K$, e $\varepsilon > 0$ segue-se que $y^T b > 0$. Agora falta mostar que $y^T A \leq 0$.

Vimos anteriormente que K é um cone, isto é, $\forall z \in K, \forall \rho > 0$, $\rho z \in K$. Consequentemente, para $z = Ae_j \in K$, com $j \in \{1, ..., n\}$ (onde e_j denota o j-ésimo vetor da base canônica do $\mathbb{R}^n$). Por conseguinte:

$$\forall \rho > 0, \;\; \rho y^T a_j = \rho y^T Ae_j = \rho y^T z = y^T \rho z \leq y^T b - \varepsilon$$

Afirmamos que, na desigualdade acima, $y^T a_j \leq 0$, pois caso contrário, bastaria tomar $\rho = \frac{y^T b - \varepsilon + 1}{y^T a_j}$, para ε suficientemente pequeno e teríamos uma contradição na última desigualdade escrita. Portanto concluímos que quando (A.1) não é solúvel, (A.2) tem solução.

■

Valendo-se das propriedades de transposição de matrizes e do lema exposto acima, podemos deduzir o seguinte resultado.

Corolário A.1.13. *Dada a matriz $A \in M_{n \times m}(\mathbb{R})$ e $b \in \mathbb{R}^m$, então um e somente um dos seguintes sistemas associados tem solução.*

$$x^T A = b^T \;\; e \;\; x^T \geq 0 \tag{A.3}$$

$$Ay \leq 0 \;\; e \;\; b^T y > 0 \tag{A.4}$$

Para nossos propósitos, vamos enunciar e demonstrar o seguinte corolário o qual será útil no decorrer do trabalho.

Corolário A.1.14. *Dada a matriz $A = (a_{ij}) \in M_{n\times m}(\mathbb{R})$ e suponha que b_j, $j = 1, 2, ..., m$, são números reais. Se cada solução $(y_1, y_2, ..., y_m)$ das desigualdades lineares homogêneas*

$$\sum_{j=1}^{m} a_{ij} y_j \geq 0 \qquad i = 1, 2, ..., n \tag{A.5}$$

é também solução da desigualdade linear homogênea

$$\sum_{j=1}^{m} b_j y_j \geq 0, \tag{A.6}$$

então existem números reais

$$x_1 \geq 0,\ x_2 \geq 0,\ ...,\ x_n \geq 0, \tag{A.7}$$

tais que

$$\sum_{i=1}^{n} x_i a_{ij} = b_j \qquad j = 1, 2, ..., m. \tag{A.8}$$

Demonstração: Utilizando a notação matricial, devido as hipóteses dadas temos que toda a solução da desigualdade $Ay \geq 0$ é também solução de $b^T y \geq 0$. Afirmamos que as desigualdades presentes em (A.4) não podem ser satisfeitas simultaneamente. Com efeito, suponhamos, por absurdo, que as desigualdades presentes (A.4) fossem satisfeitas simultaneamente. Assim, existe $y^* \in \mathbb{R}^m$ tal que: $Ay^* \leq 0$ e $b^T y^* > 0$. Consequentemente: $A(-y^*) \geq 0$ e $b^T(-y^*) < 0$. Como, $A(-y^*) \geq 0$, segue-se da hipótese dada que $b^T(-y^*) \geq 0$, o que contradiz a desigualdade $b^T(-y^*) < 0$. Portanto constatamos a veracidade da última afirmação efetuada. Agora, aplicando o corolário A.1.13 segue-se que existem $x_1 \geq 0$, $x_2 \geq 0$, ..., $x_n \geq 0$ tais que, $\sum\limits_{i=1}^{n} x_i a_{ij} = b_j$ $j = 1, 2, ..., m$.

■

A.2 Convexidade

O propósito principal desta seção é estudar as propriedades de conjuntos convexos e funções convexas os quais são necessários no desenvolvimento da teoria de programação geométrica. No que segue, nesta seção, E_N denota o espaço vetorial N-dimensional.

Teorema A.2.1. *Se A e B são subconjuntos convexos de E_N, então, a intersecção $A \cap B$ é também um subconjunto convexo de E_N.*

Demonstração: Sejam A e B são subconjuntos convexos de E_N. Suponhamos que x e y sejam pontos arbitrários, mas fixos em $A \cap B$. Então, $x, y \in A$ e também a B. Como A e B são convexos, segue-se que, $[x, y] \subset A$ e $[x, y] \subset B$. Logo, $[x, y] \subset (A \cap B)$. Portanto, $A \cap B$ é convexo. ■

Definição A.2.2. *Uma função f definida sobre um subconjunto convexo S de E_N é denominada convexa se f satisfaz a desigualdade*

$$f(\delta_1 x + \delta_2 y) \leq \delta_1 f(x) + \delta_2 f(y)$$

para cada ponto $\delta_1 x + \delta_2 y$ pertencente a um segmento de reta arbitrário contido em S. Uma função f é denominada estritamente convexa quando a desigualdade estrita

$$f(\delta_1 x + \delta_2 y) < \delta_1 f(x) + \delta_2 f(y)$$

é satisfeita para $x \neq y$ e $\delta_1, \delta_2 \neq 0$.

Teorema A.2.3. *Uma função f com derivadas parciais de segunda ordem contínuas sobre um subconjunto convexo S de E_N é convexa em S, se a função*

$$\varphi(s) = f([1 - s]x + sy), \quad 0 \leq s \leq 1$$

satisfaz a desigualdade

$$\frac{d^2\varphi}{ds^2} \geq 0, \qquad 0 \leq s \leq 1.$$

para pontos arbitrários, x e y pertencentes a S. Além disso, f é estritamente convexa se a desigualdade precede é estrita para pontos x e y distintos em S.

Demonstração: Sejam x e y pontos arbitrários, mas fixos em S e suponhamos que $\delta_1 x + \delta_2 y$ seja um ponto arbitrário no segmento de reta $[x, y]$. Então, $\delta_1 = 1 - \delta$ e $\delta_2 = \delta$ para algum $\delta \in [0, 1]$. De acordo com a fórmula de Taylor aplicada no ponto s temos as duas igualdades:

$$\varphi(0) - \varphi(s) = \left.\frac{d\varphi}{ds}\right|_s (-s) + \frac{1}{2} \left.\frac{d^2\varphi}{ds^2}\right|_{s'} s^2$$

e

$$\varphi(1) - \varphi(s) = \left.\frac{d\varphi}{ds}\right|_s (1-s) + \frac{1}{2} \left.\frac{d^2\varphi}{ds^2}\right|_{s''} (1-s)^2$$

onde s' é algum número real no intervalo $[0, s]$ e s'' é algum número real no intervalo $[s, 1]$. Portanto, multiplicando-se a primeira equação por $[1 - s]$ a segunda por s, e somando-se os resultados, obtemos:

$$\begin{aligned}[1-s]\left(\varphi(0) - \varphi(s)\right) + s\left(\varphi(1) - \varphi(s)\right) = {} & \left[\tfrac{d\varphi}{ds}\Big|_s (-s) + \tfrac{1}{2} \tfrac{d^2\varphi}{ds^2}\Big|_{s'} s^2\right](1-s) + \\ & + \left[\tfrac{d\varphi}{ds}\Big|_s (1-s) + \tfrac{1}{2} \tfrac{d^2\varphi}{ds^2}\Big|_{s''} (1-s)^2\right](s)\end{aligned}$$

Ou seja,

$$[1-s]\varphi(0) - \varphi(s) + s\varphi(1) = \frac{1}{2} \left.\frac{d^2\varphi}{ds^2}\right|_{s'} (1-s)s^2 + \frac{1}{2} \left.\frac{d^2\varphi}{ds^2}\right|_{s''} (1-s)^2 s.$$

Lembrando-se que: $\delta_1 = 1 - \delta$, $\delta_2 = \delta$, e da definição de φ. Substituímos estes dados na equação acima e obtemos:

$$\delta_1 f(x) + \delta_2 f(y) - f(\delta_1 x + \delta_2 y) = \frac{1}{2} \left.\frac{d^2\varphi}{ds^2}\right|_{s'} \delta_1 \delta_2^2 + \frac{1}{2} \left.\frac{d^2\varphi}{ds^2}\right|_{s''} \delta_1^2 \delta_2.$$

Desde que $\delta_1 \geq 0$, $\delta_2 \geq 0$ e por hipótese, $\frac{d^2\varphi}{ds^2} \geq 0$, segue-se que, $f(\delta_1 x + \delta_2 y) \leq \delta_1 f(x) + \delta_2 f(y)$. Além disso, quando $\delta_1 > 0$, $\delta_2 > 0$ e $\frac{d^2\varphi}{ds^2} > 0$, segue-se da última igualdade escrita a segunda conclusão deste teorema. ■

Teorema A.2.4. *Seja f uma função convexa definida em um conjunto convexo S. Se f é diferenciável em $x \in S$, então,*

$$f(x) + \nabla f(x) \cdot [y - x] \leqslant f(y)$$

para cada ponto $y \in S$.

Demonstração: Seja f uma função convexa definida em um conjunto convexo S. Suponhamos que f é diferenciável em $x \in S$ e consideremos $y \in S$. Sendo S um conjunto convexo, segue-se que $[x, y] \subset S$, e aplicando a definição de função convexa, obtemos:

$$f((1 - s)x + sy) \leq (1 - s)f(x) + sf(y); \quad 0 < s \leq 1.$$

Ou equivalentemente,

$$sf(x) + f((1 - s)x + sy) - f(x) \leq sf(y); \quad 0 < s \leq 1.$$

Dividindo ambos os lados da inequação por s e observando que $(1 - s)x + sy = x + s[y - x]$, temos:

$$f(x) + \frac{f(x + s[y - x]) - f(x)}{s} \leq f(y); \quad 0 < s \leq 1.$$

Por hipótese, f é diferenciável em x, logo:

$$\lim_{s \to 0^+} \frac{f\left(x + s[y - x]\right) - f(x)}{s} = \left.\frac{df\left(x + s[y - x]\right)}{ds}\right|_{s=0} = \nabla f(x) \cdot [y - x]$$

Portanto, pela última desigualdade escrita, obtemos:

$$f(x) + \nabla f(x) \cdot [y - x] \leqslant f(y).$$

■

Teorema A.2.5. *Seja h uma função monótona, não decrescente e convexa definida em E_1. Se f é uma função convexa definida em um subconjunto convexo S de E_N, então, $h \circ f$ definida em S é convexa.*

Demonstração: Sejam x e y pontos arbitrários, porém fixos, de S. Suponhamos que δ_1 e δ_2 são números reais não negativos tais $\delta_1 + \delta_2 = 1$. Como f é convexa, temos:

$$f(\delta_1 x + \delta_2 y) \leq \delta_1 f(x) + \delta_2 f(y).$$

Sendo h não decrescente, segue-se que:

$$h(f(\delta_1 x + \delta_2 y)) \leq h(\delta_1 f(x) + \delta_2 f(y)) \tag{A.9}$$

Por outro lado, h é convexa, logo:

$$h(\delta_1 f(x) + \delta_2 f(y)) \leq \delta_1 h(f(x)) + \delta_2 h(f(y)) \tag{A.10}$$

Segue-se das desigualdades (A.9) e (A.10) que:

$$h(f(\delta_1 x + \delta_2 y)) \leq \delta_1 h(f(x)) + \delta_2 h(f(y)).$$

Portanto, $h \circ f$ é convexa em S. ■

Teorema A.2.6. *Se $f_i; i = 1, 2, ..., M$ são funções convexas definidas sobre um mesmo subconjunto convexo S de E_N, então, a combinação linear $\sum_{i=1}^{M} c_i f_i$ é também uma função convexa, quando $c_i, i \in \{1, ..., M\}$ são constantes não negativas.*

Demonstração: Sejam x e y pontos arbitrários, porém fixos, de S. Suponhamos que δ_1 e δ_2 são números reais não negativos tais $\delta_1 + \delta_2 = 1$. Como f_i são funções convexas, temos:

$$f_i(\delta_1 x + \delta_2 y) \leq \delta_1 f_i(x) + \delta_2 f_i(y) \qquad i \in \{1, ..., M\}.$$

Desde que os coeficientes c_i são constantes não negativas, inferimos que:

$$c_i f_i(\delta_1 x + \delta_2 y) \leq \delta_1 c_i f_i(x) + \delta_2 c_i f_i(y) \qquad i \in \{1, ..., M\}.$$

Portanto:

$$\sum_{i=1}^{M} c_i f_i(\delta_1 x + \delta_2 y) \leq \delta_1 \left(\sum_{i=1}^{M} c_i f_i(x) \right) + \delta_2 \left(\sum_{i=1}^{M} c_i f_i(y) \right)$$

conforme desejávamos. ■

Teorema A.2.7. *Seja f é uma função convexa definida sobre um subconjunto convexo S de E_N e c um número real arbitrário. Consideremos o conjunto W, $W \subset S$, definido por: $W = \{w \in S;\ f(w) \leq c\}$. Então, W é um subconjunto convexo de S.*

Demonstração: Sejam a e b pontos arbitrários, porém fixos, de W. Suponhamos que δ_1 e δ_2 são números reais não negativos tais $\delta_1 + \delta_2 = 1$. Como f é convexa, temos:

$$f(\delta_1 a + \delta_2 b) \leq \delta_1 f(a) + \delta_2 f(b).$$

Agora, como $a, b \in W$, segue-se que:

$$\delta_1 f(a) + \delta_2 f(b) \leq \delta_1 c + \delta_2 c = (\delta_1 + \delta_2)c = c.$$

Portanto, das duas desigualdades acima, inferimos que cada ponto de $[a, b]$ está contido em W. ■

Definição A.2.8. *Uma função f definida sobre um subconjunto convexo S de E_N é denominada côncava se f satisfaz a desigualdade*

$$f(\delta_1 x + \delta_2 y) \geq \delta_1 f(x) + \delta_2 f(y)$$

para cada ponto $\delta_1 x + \delta_2 y$ pertencente a um segmento de reta arbitrário contido em S. Uma função f é denominada estritamente côncava quando a desigualdade estrita

$$f(\delta_1 x + \delta_2 y) > \delta_1 f(x) + \delta_2 f(y)$$

é satisfeita para $x \neq y$ e $\delta_1, \delta_2 \neq 0$.

Pela definição dada acima, constatamos que uma função f é côncava, se e somente se, $-f$ é convexa. Assim, temos propriedades similares as demonstradas até aqui para funções côncavas, em vez de funções convexas.

Até o momento vimos algumas propriedades de funções convexas em conjuntos convexos. Nosso estudo continuará tratando de conjuntos convexos, visto que cada função convexa e cada subconjunto convexo S presente no domínio de f origina um programa convexo, o qual consiste em minimizar a função f no conjunto S

Definição A.2.9. *Seja f uma função convexa definida em um subconjunto convexo S de E_N. Dizemos que a função $f : S \mapsto \mathbb{R}$ tem um mínimo global em $b \in S$, se $f(x) \geq f(b), \forall x \in S$.*

Deste ponto em diante, nós consideraremos apenas programas convexos os quais apresentem algum valor mínimo.

Definição A.2.10. *Seja f uma função convexa definida em um subconjunto convexo S. Dizemos que um ponto $x' \in S$ é um ponto estacionário para f, se f é diferenciável em x' e $\nabla f(x') = 0$.*

Teorema A.2.11. *Cada ponto estacionário de uma função convexa f em um conjunto convexo S é também um ponto de mínimo de f em S.*

Demonstração: Sejam x' um ponto estacionário de uma função convexa f em um conjunto S.

Pelo teorema A.2.4, temos:

$$f(x') + \nabla f(x') \cdot [y - x] \leq f(y), \ \forall y \in S.$$

Sendo x' um ponto estacionário, segue-se que:

$$f(x') \leq f(y) \;\; \forall y \in S.$$

$\blacksquare$

A.3 Algumas demonstrações

Demonstração do teorema 2.1.3: ($\Leftarrow$) Suponhamos que (z', μ') é uma solução para o problema (2). Pela primeira desigualdade presente na condição (iii) do problema (2) temos, para $\mu \geq 0$

$$L(z', \mu) - L(z', \mu') \leq 0,$$

ou seja,

$$\sum_{q=1}^{p} (\mu_q - \mu'_q) G_q(z') \leq 0.$$

Usando a condição (i), notamos que podemos escolher os coeficientes μ_q da seguinte maneira:

$$\mu_q = \begin{cases} \mu'_q, & q \neq k \\ \mu'_q + 1, & q = k \end{cases}$$

Donde vem que, $G_k(z') \leq 0,\ k = 1, 2, ..., p$. Isto mostra que a restrição do problema (1) é satisfeita por z'.

Baseado na segunda desigualdade presente em (iii) do problema (2), temos que

$$L(z, \mu') - L(z', \mu') \geq 0, \forall z \in E_N.$$

Portanto,

$$G_0(z) - G_0(z') + \sum_{k=1}^{p} \mu'_k [G_k(z) - G_k(z')] \geq 0.$$

Aplicando a condição (ii) do problema (2) temos que a última desigualdade pode ser reescrita como:

$$G_0(z) - G_0(z') + \sum_{k=1}^{p} \mu'_k G_k(z) \geq 0.$$

Logo quando z satisfaz as restrições do problema (1), temos que, $G_0(z) - G_0(z') \geq 0$, ou seja, $G_0(z) \geq G_0(z')$, sempre que z obedece as restrições $G_1(z) \leq 0$, $G_2(z) \leq 0$,...,$G_p(z) \leq 0$.

($\Rightarrow$) Suponhamos que z' seja uma solução para o problema (1) e considere o conjunto de inteiros.

$$K' = \{k \geq 1;\ G_k(z') = 0\}. \tag{A.11}$$

Por hipótese, $G_k(z)$, $k = 0, 1, 2, ..., p$ é convexa e diferenciável, consequentemente, pelo teorema A.2.4 temos:

$$G_k(z') + \nabla G_k(z') \cdot [z - z'] \leq G_k(z),\ k = 1, 2, ..., p.$$

Donde vem que,

$$\nabla G_k(z') \cdot [z - z'] \leq G_k(z),\ k \in K'$$

Lembrando-se que, por hipótese, o programa dado no problema (1) é superconsistente, temos que existe um ponto z^*, tal que,

$$G_k(z^*) < 0,\ k = 1, 2, ..., p.$$

Logo,

$$\nabla G_k(z') \cdot [z^* - z'] < 0,\ k \in K'. \tag{A.12}$$

Agora, seja z um vetor arbitrário, tal que,

$$\nabla G_k(z') \cdot [z - z'] \leq 0,\ k \in K'.$$

Combinando esta desigualdade com (A.12), temos:

$$(1 - s)\nabla G_k(z') \cdot [z - z'] + s\nabla G_k(z') \cdot [z^* - z'] < 0,\ k \in K', 0 < s \leq 1.$$

Ou equivalentemente:

$$\nabla G_k(z') \cdot [(1-s)z + sz^* - z'] < 0,\ \ k \in K', 0 < s \leq 1.$$

Para facilitar a notação, definimos:

$$x = (1-s)z + sz^*, \quad 0 < s \leq 1 \tag{A.13}$$

Deste modo, a última desigualdade escrita, torna-se

$$\nabla G_k(z') \cdot [x - z'] < 0,\ \ k \in K'. \tag{A.14}$$

Aplicando o teorema A.2.4

$$G_k(z' + t[x - z']) - G_k(z') \leq t\nabla G_k(z' + t[x - z']) \cdot [x - z']. \tag{A.15}$$

Então segue-se da relação que (A.11)

$$G_k(z' + t[x - z']) \leq t\nabla G_k(z' + t[x - z']) \cdot [x - z'], k \in K'. \tag{A.16}$$

Observe agora o seguinte: Se $k \in K'$, quando $t \to 0$, sendo as derivadas parciais contínuas, temos que:

$$\nabla G_k(z' + t[x - z']) \to \nabla G_k(z').$$

E consequentemente,

$$\nabla G_k(z' + t[x - z']) \cdot [x - z'] \to \nabla G_k(z') \cdot [x - z'].$$

Ou seja,

$$\forall \varepsilon > 0,\ \exists \delta > 0;\ 0 \leq |t| < \delta \Rightarrow |\nabla G_k(z' + t[x - z']) \cdot [x - z'] - \nabla G_k(z') \cdot [x - z']| < \varepsilon$$

Assim, se $k \in K'$, por (A.14) podemos tomar $\varepsilon = -\nabla G_k(z') \cdot [x - z']$ e teremos pela definição acima que, $\exists \delta_0 > 0$ tal que,

$$0 \leq |t| < \delta_0 \Rightarrow |\nabla G_k(z' + t[x - z']) \cdot [x - z'] - \nabla G_k(z') \cdot [x - z']| < -\nabla G_k(z') \cdot [x - z'].$$

Donde segue-se que

$$0 \leq |t| < \delta_0 \Rightarrow \nabla G_k(z' + t[x - z']) \cdot [x - z'] < 0$$

Usando a relação (A.16) temos que, para

$$0 \leq t < \delta_0 \Rightarrow G_k(z' + t[x - z']) \leq t\nabla G_k(z' + t[x - z']) \cdot [x - z'] < 0, k \in K'.$$

Por outro lado, $G_k(z)$ tem derivadas parciais contínuas para $k = 0, 1, ..., p$, disso segue-se que $G_k(z)$ é contínua em z'. Desde que,

$$G_k(z') < 0, k \notin K', \tag{A.17}$$

como consequência da continuidade, vem que,

$$\exists \delta_1 > 0;\ 0 \leq |t| < \delta_1 \Rightarrow G_k(z' + t[x - z']) < 0$$

Portanto, para $t \geq 0;\ t < \min\{\delta_0, \delta_1\}$

$$G_k(z' + t[x - z']) < 0, k = 1, 2, ..., p, \tag{A.18}$$

Por hipótese, z' é ponto de mínimo para G_0, e em (A.18) vimos que $z' + t[x - z']$ satisfaz as restrições do problema (1), para $t \geq 0;\ t < \min\{\delta_0, \delta_1\}$. Logo, $G_0(z' + t[x - z']) - G_0(z') \geq 0$ para t positivo e suficientemente pequeno.

A última desigualdade aliada a (A.15) nos permite escrever

$$\nabla G_0(z' + t[x - z']) \cdot [x - z'] \geq 0$$

para t positivo e suficientemente pequeno. Tomando o limite quando $t \to 0$, concluímos que,

$$\nabla G_0(z') \cdot [x - z'] \geq 0.$$

Por (A.13) isto equivale a:

$$\nabla G_0(z') \cdot [(1 - s)z + sz^* - z'] \geq 0,\ \ 0 < s \leq 1$$

Tomando o limite quando $s \to 0$, temos:

$$\nabla G_0(z') \cdot [z - z'] \geq 0.$$

Assim acabamos de mostrar que

$$\nabla G_0(z') \cdot [z - z'] \geq 0$$

quando z satisfaz a desigualdade

$$\nabla G_k(z') \cdot [z - z'] \leq 0, \;\; k \in K'$$

Consideremos agora o sistema:

$$\sum_{k \in K'} \mu'_k \nabla G_k(z') = -\nabla G_0(z')$$

Pelo corolário A.1.14 o sistema acima tem solução não-negativa se cada solução de $\nabla G_k(z') \cdot w \geq 0, \; k \in K'$ é também solução de $-\nabla G_0(z') \cdot w \geq 0$. Esta última implicação é facilmente constatada pelo parágrafo precedente.

Portanto, existem

$$\mu'_k \geq 0, \;\; k \in K', \tag{A.19}$$

tal que,

$$\nabla G_0(z') + \sum_{k \in K'} \mu'_k \nabla G_k(z') = 0.$$

Assim, definimos:

$$\mu'_k = 0, \;\; k \notin K', \tag{A.20}$$

concluímos que,

$$\nabla G_0(z') + \sum_{k=1}^{p} \mu'_k \nabla G_k(z') = 0. \tag{A.21}$$

Por fim, considere a função de Lagrange dada por:

$$L(z, \mu) = G_0(z) + \sum_{k=1}^{p} \mu_k G_k(z)$$

Afirmamos que (z', μ') citados acima constituem a solução para o problema (2). De fato, as relações (A.19) e (A.20) mostram que a condição (i) do problema (2) é satisfeita. Enquanto que as relações (A.11) e (A.20) provam que a condição (ii) é satisfeita.

Agora temos que verificar as desigualdades presentes em (iii). Abordaremos inicialmente a primeira delas.

Utilizando (A.11), temos que,

$$L(z', \mu) = G_0(z') + \sum_{k \notin K'} \mu_k G_k(z').$$

Por (ii), vem que, $L(z', \mu') = G_0(z')$. Valendo-se de que, $\mu \geq 0$ e
$G_k(z') < 0$, $k \notin K'$. Segue-se que, $L(z', \mu) \leq L(z', \mu')$.

Agora, pelas relações (A.19), (A.20) e pelo teorema A.2.6 segue-se que $L(z, \mu')$ é convexa em relação a z. A relação (A.21) nos mostra que z' é um ponto estacionário para $L(z, \mu')$, logo pelo teorema A.2.11, temos que, $L(z', \mu') \leq L(z, \mu')$. Com isso concluímos que realmente (z', μ') é uma solução para o problema (2).

■

Demonstração do teorema 2.1.4: Usando o teorema A.2.3, temos que a função exponencial é convexa. Aplicando o teorema A.2.4 temos que, $e^x \geq e^z + e^z(x - z)$. Usando a estratégia de redução ao absurdo, e a fórmula de Taylor para a função exponencial mostra-se que a igualdade ocorre, se e somente se, $x = z$.

Seja $y = e^z$, a inequação anterior pode ser reescrita como,

$$y + (x - \log y)y \leq e^x,$$

para qualquer x e qualquer y positivo com igualdade ocorrendo se e somente se $x = \log y$. Portanto,

$$\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} (x_i - \log y_i)\, y_i \leq \sum_{i=1}^{N} e^{x_i},$$

para cada vetor $x = (x_1, x_2, ..., x_N)$ e cada vetor positivo $y = (y_1, y_2, ..., y_N)$. A condição para que a igualdade ocorra, é a igualdade $x_i = \log y_i \quad i = 1, 2, ..., N$.

Escolhendo um vetor positivo arbitrário δ, porém fixo, nós inferimos da última desigualdade que

$$\sum_{i=1}^{N} \sigma\delta_i + \sum_{i=1}^{N} (x_i - \log(\sigma\delta_i))(\sigma\delta_i) \leq \sum_{i=1}^{N} e^{x_i}. \tag{A.22}$$

para um vetor arbitrário x e um número positivo σ. Como antes, esta desigualdade torna-se uma igualdade, se e somente se,

$$x_i = \log\delta_i + \log\sigma; \; i = 1, 2, ..., N. \tag{A.23}$$

Definimos a função auxiliar, $f : \mathbb{R}_+^* \to \mathbb{R}$, dada por,

$$f(\sigma) = \sum_{i=1}^{N} \sigma\delta_i + \sum_{i=1}^{N} (x_i - \log(\sigma\delta_i))(\sigma\delta_i).$$

a qual pode ser reescrita como,

$$f(\sigma) = \left[\sum_{i=1}^{N} \delta_i (1 + x_i - \log\delta_i)\right]\sigma - \left[\sum_{i=1}^{N} \delta_i\right]\sigma\log\sigma. \tag{A.24}$$

Com isso a desigualdade (A.22) pode ser reeditada como

$$f(\sigma) \leq \sum_{i=1}^{N} e^{x_i}, \; \sigma > 0. \tag{A.25}$$

Defina, $\psi(s) = f([1-s]u + sv), u, v > 0; 0 \leq s \leq 1$. Derivando ψ temos:

$$\psi'(s) = \left[\sum_{i=1}^{N} \delta_i (1 + x_i - \log\delta_i)\right](v-u) - \left[\sum_{i=1}^{N} \delta_i\right][(v-u) + (v-u)\log(u + s(v-u))]$$

$$\psi''(s) = -\left[\sum_{i=1}^{N} \delta_i\right]\left[\frac{(v-u)^2}{[1-s]u + sv}\right] < 0; \; u \neq v$$

Baseado na desigualdade acima, e aplicando o teorema A.2.3, temos que temos que $-f$ é convexa. Aplicando o teorema A.2.11 temos que todo o ponto estacionário, é um ponto de mínimo para $-f$. Equivalentemente, todo o ponto estacionário será um ponto de máximo para f. Como $\psi'' < 0$, temos a garantia de unicidade do ponto estacionário.

A existência do ponto estacionário, depende da resolução da equação,

$$f'(\sigma) = 0. \tag{A.26}$$

Derivando (A.24), obtemos:

$$f'(\sigma) = \left[\sum_{i=1}^{N} \delta_i \left(1 + x_i - \log \delta_i\right)\right] - \left[\sum_{i=1}^{N} \delta_i\right] \left[\log(\sigma) + 1\right].$$

,ou seja,

$$f'(\sigma) = \left[\sum_{i=1}^{N} \delta_i \left(x_i - \log \delta_i\right)\right] - \left[\sum_{i=1}^{N} \delta_i\right] \left[\log \sigma\right]. \tag{A.27}$$

Resolvendo a equação (A.26), temos que o único ponto de máximo σ' é dado por:

$$\log(\sigma') = \frac{\left[\sum\limits_{i=1}^{N} \delta_i \left(x_i - \log \delta_i\right)\right]}{\sum\limits_{i=1}^{N} \delta_i} \tag{A.28}$$

Substituindo σ' em (A.25) e usando (A.24) e (A.28), obtemos:

$$\left[\sum_{i=1}^{N} \delta_i\right] \sigma' \leq \sum_{i=1}^{N} e^{x_i}.$$

E como a função logarítmica é crescente, temos:

$$\log\left(\sum_{i=1}^{N} \delta_i\right) + \log\left(\sigma'\right) \leq \log\left(\sum_{i=1}^{N} e^{x_i}\right).$$

Assim, substituindo a expressão $\log(\sigma')$ dada em (A.28) na desigualdade anterior segue-se que:

$$\sum_{i=1}^{N} x_i\delta_i \leq \left(\sum_{i=1}^{N} \delta_i\right) \log\left(\sum_{i=1}^{N} e^{x_i}\right) + \sum_{i=1}^{N} \delta_i \log\delta_i - \left(\sum_{i=1}^{N} \delta_i\right) \log\left(\sum_{i=1}^{N} \delta_i\right).$$

Afirmamos que a igualdade ocorre se, e somente se,

$$x_i = \log \delta_i + \frac{\sum\limits_{j=1}^{N} \delta_j \left(x_j - \log \delta_j \right)}{\sum\limits_{k=1}^{N} \delta_k}.$$

De fato: $(\Leftarrow)$ $x_i = \log \delta_i + \frac{\sum\limits_{j=1}^{N} \delta_j (x_j - \log \delta_j)}{\sum\limits_{k=1}^{N} \delta_k}$; $i = 1, ..., N$, por (A.28), vem que, $x_i = \log \delta_i + \log \sigma'$. Assim: $e^{x_i} = \delta_i \sigma', i = 1, 2, ..., N$.

Logo, $\sum\limits_{i=1}^{N} e^{x_i} = \sum\limits_{i=1}^{N} \delta_i \sigma'$, e consequentemente,

$$\log \sum_{i=1}^{N} e^{x_i} = \log \sum_{i=1}^{N} \delta_i + \log \sigma'.$$

Substituindo o valor de $\log \sigma'$ obtemos:

$$\log \sum_{i=1}^{N} e^{x_i} = \log \sum_{i=1}^{N} \delta_i + \frac{\sum\limits_{j=1}^{N} \delta_j \left(x_j - \log \delta_j \right)}{\sum\limits_{k=1}^{N} \delta_k}.$$

Donde vem que fazendo operações aritméticas, obtemos:

$$\sum_{i=1}^{N} x_i \delta_i = \left(\sum_{i=1}^{N} \delta_i \right) \log \left(\sum_{i=1}^{N} e^{x_i} \right) + \sum_{i=1}^{N} \delta_i \log \delta_i - \left(\sum_{i=1}^{N} \delta_i \right) \log \left(\sum_{i=1}^{N} \delta_i \right).$$

$(\Rightarrow)$

$$\sum_{i=1}^{N} x_i \delta_i = \left(\sum_{i=1}^{N} \delta_i \right) \log \left(\sum_{i=1}^{N} e^{x_i} \right) + \sum_{i=1}^{N} \delta_i \log \delta_i - \left(\sum_{i=1}^{N} \delta_i \right) \log \left(\sum_{i=1}^{N} \delta_i \right).$$

$$\Leftrightarrow$$

$$\sum_{i=1}^{N} [x_i - \log \delta_i] \delta_i + \left(\sum_{i=1}^{N} \delta_i \right) \log \left(\sum_{i=1}^{N} \delta_i \right) = \left(\sum_{i=1}^{N} \delta_i \right) \log \left(\sum_{i=1}^{N} e^{x_i} \right).$$

$$\Leftrightarrow$$

$$\frac{\sum\limits_{i=1}^{N} [x_i - \log \delta_i] \delta_i}{\left(\sum\limits_{i=1}^{N} \delta_i \right)} + \log \left(\sum_{i=1}^{N} \delta_i \right) = \log \left(\sum_{i=1}^{N} e^{x_i} \right). \tag{A.29}$$

Vimos anteriormente, que fazendo,

$$\log(\sigma') = \frac{\left[\sum\limits_{i=1}^{N} \delta_i \left(x_i - \log \delta_i\right)\right]}{\sum\limits_{i=1}^{N} \delta_i},$$

temos, $f(\sigma') = \sigma' \sum\limits_{i=1}^{N} \delta_i$. Por outro lado, pela equação (A.29), temos,

$$\log \sigma' + \log \left(\sum_{i=1}^{N} \delta_i\right) = \log \left(\sum_{i=1}^{N} e^{x_i}\right)$$

Consequentemente,

$$f(\sigma') = \sigma' \sum_{i=1}^{N} \delta_i = \left(\sum_{i=1}^{N} e^{x_i}\right).$$

De acordo com (A.23), temos que $x_i = \log \delta_i + \log \sigma'; i = 1, ..., N$.

$$x_i = \log \delta_i + \frac{\sum\limits_{j=1}^{N} \delta_j \left(x_j - \log \delta_j\right)}{\sum\limits_{k=1}^{N} \delta_k}; i = 1, ..., N.$$

Ou equivalentemente, $x_i = \log \delta_i - \log B; i = 1, ..., N$. Para algum número positivo B. Assim provamos a afirmação proferida, e o teorema para um vetor δ com todas as componentes positivas.

Caso todas as componentes de δ são nulas, a desigualdade é claramente satisfeita, pois ambos os lados da desigualdade são nulos.

Caso nem todas as coordenadas de δ sejam positivas, podemos empregar uma mesma reordenação nos vetores δ e x chegando a um vetor $\widetilde{\delta}$ e $\widetilde{x}$, de modo que, $\widetilde{\delta_i} > 0; i = 1, 2, ..., s$ e $\widetilde{\delta_i} = 0; i = s + 1, ..., N$. Note que tal reordenação não altera a desigualdade a ser demonstrada. Com isso, sem perda de generalidade, podemos supor que,

$$\delta_i > 0, \quad i = 1, 2, ..., s. \tag{A.30}$$

$$\delta_i = 0, \quad i = s + 1, ..., N. \tag{A.31}$$

onde $1 \leq s \leq N$. A partir do que foi provado anteriormente,

$$\sum_{i=1}^{s} x_i\delta_i \leq \left(\sum_{i=1}^{s} \delta_i\right) \log \left(\sum_{i=1}^{s} e^{x_i}\right) + \sum_{i=1}^{s} \delta_i \log\delta_i - \left(\sum_{i=1}^{s} \delta_i\right) \log \left(\sum_{i=1}^{s} \delta_i\right),$$

por (A.31), temos:

$$\sum_{i=1}^{N} x_i\delta_i \leq \left(\sum_{i=1}^{N} \delta_i\right) \log \left(\sum_{i=1}^{s} e^{x_i}\right) + \sum_{i=1}^{N} \delta_i \log\delta_i - \left(\sum_{i=1}^{N} \delta_i\right) \log \left(\sum_{i=1}^{N} \delta_i\right).$$

Agora, como $e^{x_i} > 0; i = s+1, ..., N$, e a função logaritmo é não decrescente, temos que,

$$\sum_{i=1}^{N} x_i\delta_i < \left(\sum_{i=1}^{N} \delta_i\right) \log \left(\sum_{i=1}^{N} e^{x_i}\right) + \sum_{i=1}^{N} \delta_i \log\delta_i - \left(\sum_{i=1}^{N} \delta_i\right) \log \left(\sum_{i=1}^{N} \delta_i\right).$$

A prova esta completa, pois não existe B, tal que, $\delta_i = Be^{x_i}; i = 1, ..., N$. ■

Referências Bibliográficas

[1] R. J. Duffin; E. L. Peterson; C. Zener, *Geometric Programming - Theory and application* . John Wiley and Sons,Inc, New York, 1967.

[2] R. A. Horn; C. R. Johnson, *Topics in Matrix Analysis.* Cambridge University Press, 1994.

[3] M. Ibragimov, *A method of calculating the spectral radius of a nonnegative matrix and its applications.* Economic Theory 17: 467-480, 2001.

[4] E. L. Lima, *Álgebra linear* . IMPA, Rio de Janeiro, 1998.

[5] E. L. Lima, *Análise Real - Volume I* . IMPA, Rio de Janeiro, 1997.

[6] E. L. Lima, *Curso de análise-volume I* . IMPA, Rio de Janeiro, 2008.

[7] E. L. Lima, *Curso de análise-volume II* . IMPA, Rio de Janeiro, 2011.

[8] E. L. Lima, *Espaços Métricos* . IMPA, Rio de Janeiro, 2005.

[9] C. D. Meyer, *Matrix analisys and applied linear algebra.* SIAM, Philadelphia, 2000.

[10] M. Parry; M. Pollicott, *Zeta functions and the periodic orbit structure of hyperbolic dynamics.* Astérisque Vol 187-188, 1990

[11] M. Pollicott; M. Yuri, *Dynamical systems and ergodic theory*. Camb. Univ. Press, Cambridge, 1998.

[12] M. A. G. Ruggiero; V. L. R. Lopes, *Cálculo numérico: aspectos teóricos e computacionais*. 2^a ed., Makron Books, São Paulo, 1996.